THE UNIVERSITY OF WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/69774

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

Nutrient Translocation to Offspring in Plants

Perry James Dominic Bateman

A thesis submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy in Life Sciences

University of Warwick, School of Life Sciences

September 2014

Acknowledgments

While in many ways a PhD is quite a solitary journey, there are many people that made invaluable contributions without which I could not possibly have succeeded. I would like to thank these people here for their support and encouragement that kept me motivated along the way.

I would first like to thank my supervisors, Dr Jose Gutierrez-Marcos and Wyatt Paul for all the generous help, encouragement and technical advice throughout my research. I would particularly like to thank Jose for giving me my first job in science after my undergraduate degree and encouraging me to apply for my PhD project. I would also like to thank Biogemma for all their assistance in generation of transgenic lines, statistical analyses and financial support. Thanks also to Prof. Shinya Ohki for invaluable assistance in protein expression and structural work. I would like to thank all the past and present members of the Marcos research group for support, camaraderie and encouragement, and particularly Jenny Goodman, Ranjith Papareddy and Julius Durr for their technical assistance.

Next I would like to thank my parents for their continuous support and encouragement not just during my PhD, but throughout my education. This support allowed me to pursue my passion for science at university and enjoy all the other fantastic parts of university life. My amazing and now extended family have also played a huge part in my life at Warwick, setting the bar high for achievement and offering fantastic support and great company which kept me going.

My friends both at Warwick and from home have made my time at university immensely enjoyable, a huge factor in why I stayed so long! Particular thanks to Dave, Charlie and Matt who went through the PhD journey with me, keeping me sane with copious amounts of gym time and fried chicken.

Finally, my biggest thanks go to my wife Elle who has supported and encouraged me throughout. Thank you Elle for being there for me and for inspiring me to do my best, you have helped me through the tough times and laughed with me through the fun. Thank you for being my best friend, I could not have done it without you.

Declaration

The work presented in this thesis is original, and has not been published or presented for any other degree.

Contents

1 Introduction **1**

1.1 General Overview 2

1.2 Plant reproductive development 4

1.2.1 Double fertilisation - Developmental divergence 4

1.2.2 Plant Embryogenesis - Coordinated biological choreography . . 4

1.2.3 The Embryo Suspensor - An umbilical role 9

1.2.4 Endosperm Development - A tale of altruistic sacrifice 10

1.2.5 Differentiation of the Endosperm Tissues - Specialism defined . 11

1.2.6 Aleurone - The cells of the front-line 12

1.2.7 BETL - The nutrient trading-floor 13

1.2.8 The Embryo Surrounding Region - To protect and serve 15

1.3 Small Cysteine-Rich Proteins in Plants - Jack of all trades 20

1.3.1 Reproductive roles for CRPs - From speed-dating to wedlock . . 23

1.3.2 CRP structure - Constrained disorder 27

1.4 Project Aims 29

2 Materials and Methods **30**

2.1 General Plant Materials and Methods 31

2.1.1 *Zea mays* plant growth 31

2.1.2 *Nicotiana benthamiana* plant growth 31

2.1.3 *Arabidopsis thaliana* plant growth 31

2.1.4 Transgenic maize plantlet growth 32
2.1.5 Maize controlled pollination 32
2.1.6 *Arabidopsis* controlled pollination 32
2.1.7 Generation of pro*Mlp1*:GUS transgenic maize lines 33
2.1.8 Generation of MLP1-GFP transgenic maize lines 33
2.1.9 Generation of maize pro*Mlp1*:RNAi transgenic lines 34
2.1.10 Transformation of *Agrobacterium* by electroporation 34
2.1.11 *Agrobacterium tumefaciens* mediated transient expression in Tobacco leaves 35
2.1.12 Transformation of *Arabidopsis thaliana* 36
2.1.13 BASTA selection of transgenic *Arabidopsis thaliana* 36
2.1.14 *Arabidopsis thaliana* seed sterilisation 36
2.1.15 Transgene insertion quantification 37
2.2 General Molecular Materials and Methods 37
2.2.1 RNA extraction from maize kernels 37
2.2.2 RNA extraction by hot-phenol method 38
2.2.3 cDNA synthesis 38
2.2.4 qRT-PCR primer design and verification 39
2.2.5 Semi-Quantitiative PCR 40
2.2.6 Quantitative Real-Time PCR 40
2.2.7 Maize Genomic DNA extraction 41
2.2.8 Maize genomic DNA PCR reactions 41

2.2.9 Genotyping transgenic seeds by GFP reporter gene 42
2.2.10 Genotyping transgenic plants by molecular methods 42
2.2.11 Generation of endosperm specific cDNA libraries 42
2.3 Cloning Methods 43
2.3.1 Synthetic gene design 43
2.3.2 Cloning synthetic genes into pET29a expression vector 44
2.3.3 Transformation of DH5α competent *E.coli* cells 44
2.3.4 Identification of positive transformants by colony PCR 44
2.3.5 Sequencing of positive transformants 45
2.3.6 Transformation into Rosetta 2 DE3 cells 45
2.3.7 Gateway Cloning 45
2.3.8 *At*ESF1.3/*Zm*MLP1 domain substitution constructs 46
2.4 Biochemical Methods 47
2.4.1 Protein expression in Rosetta *E.coli* 47
2.4.2 Protein extraction from maize seeds 47
2.4.3 Protein extraction by TCA NaDOC precipitation 48
2.4.4 Protein extraction by acetone precipitation 48
2.4.5 Protein quantification 48
2.4.6 SDS-PAGE of proteins 49
2.4.7 Tricine-SDS PAGE of proteins 49
2.4.8 Protein purification by metal affinity 49

2.4.9 Protein purification by Immobilised Metal Affinity Chromatography (IMAC) . . . 50
2.4.10 Protein identification by Mass Spectrometry . . . 50
2.4.11 Antibody Production . . . 51
2.4.12 Enumeration of disulphide bonds in protein by Ellman's reaction 51
2.4.13 MLP1 secondary structure assessment by circular dichroism . . 51
2.4.14 pro*Mlp1*-GUS expression and detection . . . 52
2.4.15 Immunodetection of proteins . . . 52
2.4.16 Extraction of apoplastic and total protein from *N.benthamiana* leaves . . . 53
2.4.17 Immunoprecipitation of GFP fusion proteins . . . 53
2.4.18 Protein expression in BY2 cell suspension culture . . . 54
2.4.19 Protein purification and structural NMR . . . 54
2.5 Histological and Microscopic Methods . . . 54
2.5.1 Maize kernel fixation . . . 54
2.5.2 Tissue sectioning . . . 55
2.5.3 Immunohistochemistry . . . 55
2.5.4 *In situ* hybridisation . . . 56
2.5.5 Laser scanning confocal microscopy of GFP fusion proteins . . . 58
2.5.6 Epifluorescence microscopy of GFP fusion proteins . . . 59
2.5.7 Internal seed morphology analysis . . . 59
2.5.8 Histological staining . . . 60

2.6 *In silico* Methods 60
2.6.1 Amino acid sequence alignments 60
2.6.2 Identification of *At*ESF3/*Zm*MLP1 structural loops 60
2.6.3 Promoter Region Sequence Analysis 61
2.6.4 Analysis of protein biophysical properties 61
2.7 Phenotype analysis methods 61
2.7.1 Measurement of dry seed area 61
2.7.2 Dry seed weight measurements 62
2.7.3 Measurement of starch proportional area 62
2.7.4 Dimensional analysis of developing maize embryos 62
2.7.5 Measurement of *Arabidopsis* suspensor length 63
2.8 Plant tissue culture 63
2.8.1 Calli induction and maintenance 63
2.8.2 Liquid plant cell suspension 63
2.8.3 Expression induction with dexamethasone (DEX) 64
2.9 Statistical Methods 64
2.9.1 Welches Students t-test 64
2.9.2 ANOVA multiple comparisons test 64
2.9.3 Chi-squared test 64

3 Discovery and Molecular Characterisation of the MEG-Like Protein 1 Gene 66

3.1 Introduction 67

3.2 Identification of MLP1 - a novel, small cysteine-rich protein 70

3.3 Expression analysis of *ZmMlp1* 73

3.3.1 The *ZmMlp1* promoter is spatially and temporally specific to the ESR in the early seed 74

3.3.2 The *ZmMlp1* gene expression is temporally regulated 76

3.4 MLP1 Protein Expression Analysis 79

3.4.1 Expression and purification of MLP1 79

3.5 *Zm*MLP1 Antibody Production and Analysis 82

3.6 *Zm*MLP1 is expressed in the central cell and ESR 84

3.7 *Zm*MLP1 contains a secretory signal peptide 92

3.8 Discussion 96

4 Functional Characterisation of *ZmMlp1* 101

4.1 Introduction 102

4.2 Functional investigation of *Zm*MLP1 by RNA interference 105

4.3 Phenotypic characterisation of *ZmMlp1* RNAi lines 108

4.3.1 *Zm*MLP1 is required for normal mature seed morphology 108

4.3.2 *Zm*MLP1 is involved in early embryo pattern formation 116

4.3.3 Molecular analysis of the DEP phenotype 119

4.4 Genetic complementation of *ZmMlp1* RNAi by *Zm*MLP1-GFP 130

4.5 Discussion 141

5 Structure-Function Analysis of *Zm*MLP1 **147**

5.1 Introduction 148

5.2 *In silico* structural analysis of *Zm*MLP1 152

5.3 Expression of active *Zm*MLP1 153

5.4 Solving the 3D structure of *Zm*MLP1 158

5.5 Interspecific structure-functional analysis of CRPs 164

5.6 Discussion 177

6 Conclusions and Further Work **180**

6.1 *Zm*MLP1 is an evolutionarily conserved ESR factor 181

6.2 *Zm*MLP1: Structure-function relationships 183

6.3 *Zm*MLP1 - A maternally required endosperm factor 185

7 Appendix **188**

7.1 Synthetic gene constructs 188

7.2 Vector Maps 189

7.3 Sequence Alignments 195

7.4 PCR primer sequences 195

References **199**

List of Figures

1.1 2011 global food production statistics 3

1.2 Developmental stages of the maize seed 7

1.3 Developmental stages of the maize embryo 8

3.1 Schematic comparison of maize and sorghum chromosomes 68

3.2 Structure and sequence of the *ZmMlp1* gene and protein 72

3.3 Electronic Fluorescent Pictograph map of *ZmMlp1* 73

3.4 Spatial and temporal activity of the *ZmMlp1* promoter 75

3.5 Temporal expression profile of the ZmMlp1 gene 77

3.6 *In situ* hybridisation of *ZmMlp1* 78

3.7 Expression and purification of *Zm*MLP1 from *E.coli* 81

3.8 Evaluation of custom generated MLP1 antibody by Western blotting . 84

3.9 Cellular localisation of the MLP1 protein 86

3.10 Microscopic analysis of MLP1-GFP expressed in unfertilised maize ovules 88

3.11 Microscopic analysis of MLP1-GFP expressed in developing maize seeds at 4 DAP 89

3.12 Microscopic analysis of MLP1-GFP expressed in developing maize seeds at 6 DAP 89

3.13 Microscopic analysis of MLP1-GFP expressed in developing maize seed at 8 DAP 90

3.14 Microscopic analysis of MLP1-GFP expressed in developing maize seed at 8 DAP 91

3.15 Detection of *Zm*MLP1-GFP from isolated endosperms at 8 DAP by Western blot 92

3.16 Signal peptide cleavage site in *Zm*MLP1 95

4.1 RNAi based down-regulation of ZmMlp1 expression 107

4.2 Dry seed analysis of pro*Mlp1*-RNAi ears 109

4.3 Dry seed weight and size analysis of pro*Mlp1*-RNAi lines 111

4.4 Dry seed cross-section and starch content analysis of pro*Mlp1*-RNAi lines 114

4.5 In situ hybridisation analysis of MLP1 RNAi lines at 6 DAP 117

4.6 Dimensional analysis of proembryos in developing *ZmMlp1* RNAi seeds at 6 DAP 119

4.7 Differential gene expression analysis of *ZmMlp1*-RNAi plants by RNA-seq 122

4.8 Expression analysis of endosperm compartment specific genes by qRT-PCR 128

4.9 Starch content of reciprocal crosses to test genetic complementation of pro*Mlp1*:RNAi by *Zm*MLP1-GFP at 15 DAP 132

4.10 Starch content and seed size of reciprocal crosses to test genetic complementation of pro*Mlp1*:RNAi by *Zm*MLP1-GFP in mature seeds 133

4.11 Histological analysis of genetic complementation of *ZmMlp1* RNAi by *Zm*MLP1-GFP 136

4.12 Expression profile of reciprocal crosses between A188 and *ZmMlp1*-GFP lines at different developmental stages by qRT-PCR 140

5.1 Developmental stages of the *Arabidopsis* embryo 152

5.2 General and predicted secondary structure of *Zm*MLP1 154

5.3 Circular Dichroism spectrum fit for recombinant *Zm*MLP1 156

5.4 Plant based expression of MLP1 158

5.5 NMR structural conformation analysis of *Zm*MLP1 160

5.6 Structural and electrostatic surface representations of *Zm*MLP1 162

5.7 Schematic alignment of *Zm*MLP1 homologues across species 164

5.8 Identification of *Zm*MLP1 structural loops 166

5.9 Generation of chimeric *Zm*MLP1-ESF1.3 proteins 169

5.10 Phenotypic analysis of *Arabidopsis* expressing chimeric proteins 170

5.11 Effect of expression of *At*ESF1.3-*Zm*MLP1 substitution constructs on suspensor length 172

5.12 Homology based models of the chimeric *At*ESF1.3-Zm*MLP1* substitution constructs 175

7.1 Schematic representation of the MLP1-GUS fusion construct used to drive GUS expression in transgenic maize. 188

7.2 Schematic representation of the synthetic construct used to express His-tagged MLP1 in *E.coli.* 188

7.3 Schematic representation of the synthetic construct used to express His-tagged MLP1 in *E.coli.* 188

7.4 Schematic representation of the construct used to drive expression of *Zm*MLP1-RNAi 188

7.5 Schematic representation of the synthetic construct used to express *Zm*MLP1 in tobacco 189

7.6 Vector map of pBIOS 1865 used to express pro*Mlp1*:GUS construct in transgenic maize plants. 189

7.7 Vector map of pBIOS 3105 used to express MLP1:GFP construct, driven by pro*Mlp1* in transgenic maize plants. 190

7.8 Vector map of pEARLEYGATE 100 used to express MLP1:GFP construct, driven by proCaMV 35S in *N.Benthamiana*. 191

7.9 Vector map of pBIOS 2755 used to express MLP1 RNAi construct, driven by pro*Mlp1* in maize. 192

7.10 Vector map of pJOD28-SCV used to express pro*ZmBETL*1:GUS in wheat. 193

7.11 Vector map of pJOD29-SCV used to express pro*ZmBETL*9:GUS in wheat. 194

7.12 MLP1 vs MEG1 DNA sequence alignment 195

7.13 MLP1 vs synthetic MLP1 DNA sequence alignmentnMLP1 195

7.14 The adventures of Marvin the Mad Maize. 198

List of Tables

1.1 Plant CRPs and their putative functions in different plant species. Adapted from (Marshall *et al.* , 2011). 26

3.1 Peptide fragments identified by nanoLC ESI MS/MS after tryptic digestion of SDS-PAGE protein band. Scores and expected values calculated from Mascot protein identification software. * indicates a significant detection according to an expected value <0.05. 83

4.1 One-way ANOVA multiple comparison test (Sidak) of mean dry seed weight in T02297 (MLP1 RNAi) lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene. 112

4.2 One-way ANOVA multiple comparison test (Sidak) of mean dry seed area in T02297 (MLP1 RNAi) lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene. 112

4.3 One-way ANOVA multiple comparison test (Sidak) of seed endosperm proportion in RNAi lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene. 115

4.4 One-way ANOVA multiple comparison test (Sidak) of seed cross-sectional area in RNAi lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene. 115

4.5 Two-tailed Student's t-test of embryo dimensions in RNAi (T02297) lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants. 120

4.6 Putative functions for selected down-regulated genes as a consequence of *ZmMLP1* RNAi. 124

4.7 Putative functions for selected up-regulated genes as a consequence of *ZmMLP1* RNAi. 125

4.8 Students t-test (Welches) of gene expression levels for compartment specific genes in developing RNAi seeds (Line T02297). (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene. 129

4.9 One-way ANOVA multiple comparison test (Sidak) of proportional starch area from reciprocal and control crosses at 15 DAP. (RNAi = Line T02297, MLP1-GFP = Line T02492). n = number of kernels pooled from 3 biological replicates. 134

4.10 One-way ANOVA multiple comparison test (Sidak) of mature seed cross-sectional area from reciprocal and control crosses.(RNAi = Line T02297, MLP1-GFP = Line T02492). n = number of kernels pooled from 3 biological replicates. 134

4.11 One-way ANOVA multiple comparison test (Sidak) of embryo suspensor dimensions at 6 DAP from reciprocal and control crosses. (RNAi = Line T02297, MLP1-GFP = Line T02492). n = number of kernels pooled from 3 biological replicates. 138

4.12 One-way ANOVA multiple comparison test (Sidak) of embryo proper and total proembryo dimensions at 6 DAP from reciprocal and control crosses. (RNAi = Line T02297, MLP1-GFP = Line T02492). n = number of kernels pooled from 3 biological replicates. 139

5.1 Predicted secondary structure fractions calculated from experimental data submitted to the DichroWeb server using the CDSSTR method and the SP175 reference database. Normalised root-mean-square deviation (NRMSD) = 0 163

5.2 Predicted phosphorylation sites in the MLP1 amino acid sequence as predicted by the NetPhos 2.0 web based server (Blom *et al.* , 1999) . . 163

5.3 One-way ANOVA multiple comparison test (Dunnett's) of suspensor lengths across different substitution lines and control crosses. 173

7.1 Genotyping PCR primer sequences 196

7.2 qRT-PCR primer sequences. Introns marked with * 196

7.3 Gateway cloning primer sequences. attB sequences in lower case 197

List of Abbreviations

At	*Arabidopsis thaliana*
bp	base pair
cDNA	complementary DNA
°C	degrees Celcius
cm	centimetres
CRP	Cysteine Rich Protein
CRD	Cysteine Rich Domain
BETL	Basal Endosperm Transfer Layer
BSA	Bovine Serum Albumin
DAP	Days After Pollination
dNTPs	Deoxynucleotide Triphosphates
EDTA	Ethylenediaminetetra-acetic Acid
ESR	Embryo Surrounding Region
g	grams
GFP	Green Fluorescent Protein
IMAC	Immobilised Metal Affinity Chromatography
L	Litres

LSCM	Laser Scanning Confocal Microscopy
LB	Luria Broth
µg	micrograms
µl	microlitres
µm	micrometres
µM	micromolar
mM	millimolar
M	Molar
mg	milligrams
min	minutes
mL	millilitres
MLP1	Meg Like Protein 1
mRNA	messenger RNA
MS	Murashige and Skoog medium
ng	nanograms
nm	nanometre
NMR	Nuclear Magnetic Resonance
PCR	Polymerase Chain Reaction
RNAi	Ribonucleic Acid interference
rpm	revolutions per minute
sec	seconds

SDS	Sodium dodecyl sulphate
qRT-PCR	Quantitative Real-Time PCR
WT	Wild type
Zm	*Zea mays*

Abstract

A key question in biology is how the transfer of nutrients from mother to offspring is regulated. In mammals, the translocation of nutrients to the embryo is mediated by the placenta, while in flowering plants is governed by the seed endosperm. Current evidence shows that in plants, nutrient transfer from maternal tissues to developing embryos is regulated at two levels: Initially, from maternal integuments to endosperm through a layer of specialised transfer cells and secondly from endosperm to embryo through the embryo-surrounding cells (ESR). However, a remaining question is: how is the translocation of nutrients between these two layers coordinated?

The aim of this work was to investigate small cysteine-rich proteins (CRP) and their involvement in the reproductive development of cereal crops. This was carried out by characterisation of *ZmMlp1*, a novel plant CRP gene, using promoter reporters, spatio-temporal gene expression profiling, biochemical protein characterisation and localisation. The function of *ZmMlp1* was then investigated by knocking down its expression in maize through RNA interference and subsequent morphological and molecular phenotypic characterisation. The 3D structure of *Zm*MLP1 was resolved by solution state NMR, allowing investigation into key structural components and how they contribute to the functionality of *Zm*MLP1.

The characterisation of *Zm*MLP1 demonstrated a novel expression pattern in maize, since *Zm*MLP1 was shown to be a secreted protein expressed in both the central cell of the ovule and the ESR of the developing endosperm between 0 and 10 DAP. Functionally, *Zm*MLP1 was shown to play a role in embryonic development, demonstrated by a defective embryo patterning phenotype displayed by the RNAi transgenics. Furthermore, seed weight and size was shown to be significantly reduced in *ZmMlp1*-RNAi lines, linking *ZmMlp1* to crop yield. The structure of *Zm*MLP1 was shown to contain a knottin-like core, stabilised by 4 disulphide bonds which support disordered sequence loops that form key sections of the molecular surface. Sequence chimeras formed from *Zm*MLP1 and *At*ESF1.3 induced a known suspensor phenotype in *Arabidopsis*, demonstrating that specific structural loops are critical to CRP functionality.

Collectively these studies have furthered the understanding of nutrient translocation between maternal tissue to offspring and how CRPs are involved in these plant processes.

1 Introduction

1.1 General Overview

As the Earth's population climbs towards 8 billion, the importance of global food security has never been more apparent. The demand for food is predicted to increase by 50% in the next 20 years, with very little additional arable land (Beddington, 2009; Jones, 2011; Cook *et al.* , 2011; Godfray *et al.* , 2010). One strategy to meet this demand is a biotechnological approach to increase yield in existing food crops (Jones, 2011). This allows a faster and more targeted approach to the introduction of new traits compared to traditional plant breeding (Bruce, 2012). Nearly half of the global crop production output for 2011 was made up of cereal crops (Figure 1.1 (a)). If the cereal crop production is then broken down further, we can see that maize is the most abundantly produced cereal in the world, making it an incredibly important crop for both global food security and the economy (Figure 1.1 (b)). This high demand for maize makes it a logical model organism for research into the overall developmental processes in the plant and how these might be exploited to increase yield potential. A more specific goal is to increase the nutritional sink strength of cereal crops, to boost the dietary value of individual seeds. This leads us to an important biological question: Which molecular mechanisms are involved in determining nutritional value, and how can these be manipulated to improve nutritional quality? One of the areas of research to this aim is to manipulate the nutrient transfer mechanisms that exist between the parent plant and the seed. This offers potential to shift the balance of nutrient allocation and hence increase yield.

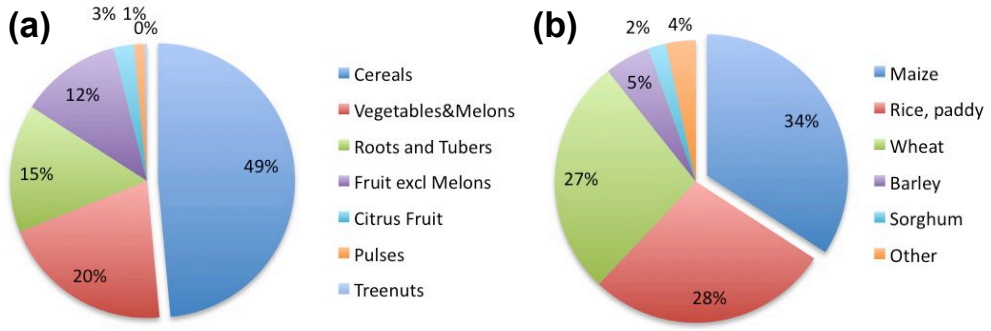

Figure 1.1: 2011 global food production statistics as published by the Food And Agriculture Organization Of The United Nations (FAO). (a) Food crop production figures grouped by crop type. (b) Cereal production figures grouped by species. All figures represented as a percentage of the total gross production value in tonnes [FAOSTAT].

1.2 Plant reproductive development

1.2.1 Double fertilisation - Developmental divergence

Angiosperms (including maize) have developed a unique and complex method of reproduction involving the process of double fertilisation. Double fertilisation was first co-discovered by Navaschin in 1898 (Kordyum, 2008) and Guignard in 1899 (Guignard, 1899) and involves the simultaneous fusion of two female and two male gametes. Pollen recognition occurs at the stigma on the pistil and subsequent pollen tube growth initiates, guided by maternal factors in the female tissue towards the female gametophyte. The female gametophyte in most cases, consists of an eight-nucleate embryo sac comprising of an egg cell, two synergid cells, three antipodal cells and a central cell containing two polar nuclei (Williams & Friedman, 2002) (Figure 1.2). The synergids in the female gametophyte attract the pollen tube, resulting in direct interaction followed by growth arrest. The two male gametes (sperm cells) are then released in the degenerated synergid followed by migration of the male gametes to the female gametes, where recognition and fusion occurs. The female gametes are the haploid egg cell and the diploid central cell in the embryo sac which after gamete fusion, developmentally diverge to become the diploid zygote and the triploid endosperm respectively (Figure 1.2). The fusion of the genetic material during karyogamy then allows the cell-cycle to restart and zygotic development to occur (Berger, 2008; Berger *et al.* , 2008).

1.2.2 Plant Embryogenesis - Coordinated biological choreography

The mature embryo is the earliest form of the plant sporophytic stage, held in stasis until germination when it develops into a new plant. To reach this point though, the embryo must pass through a concerted set of complex developmental stages, carrying a single celled zygote to a fully differentiated structure, containing primordia for the

first leaves and roots of the future seedling (Figure 1.3) (Hajduch *et al.* , 2005). The developmental stages of this process vary considerably between different plant species, this review however, will focus on the crop plant and model organism *zea mays*. The fertilisation of the first female gamete, the egg cell, by one of the male sperm cells marks the transition into the zygotic stage and phase 1 of overall embryogenesis. Interestingly, the egg cell contents are spatially polarised towards the micropylar end of the embryo sac, a polarisation which is rapidly inverted towards the antipodal end within 16 hours after fertilisation (Van Lammeren, 1986b; Mòl *et al.* , 1994).

Similarly to mammalian systems, fertilisation of the egg cell is immediately followed by a rapid and transient pulse in intracellular Ca^{2+} concentration, an event which is thought to direct the reversal of the zygote polarisation (Digonnet *et al.* , 1997; Santella *et al.* , 2004). However, in contrast to the mammalian mechanism, the $[Ca^{2+}]$ increase is due to an influx of extracellular Ca^{2+}through calcium channels, originating at the point of sperm penetration and spreading throughout the cell (Antoine *et al.* , 2000). In maize, experiments using extracellular vibrating probes and fluorescent Ca^{2+} probes have demonstrated this Ca^{2+}flux in reaction to egg cell fertilisation, resulting in egg cell activation (Antoine *et al.* , 2001). One of the processes of egg cell activation is the rapid deposition of a cell wall around the egg cell protoplast (Tirlapur *et al.* , 1995). This process is thought to be similar to the cortical reactions in the animal kingdom that protect the egg cell against polyspermy. In *Arabidopsis*, this Ca^{2+} pulse has recently been characterised in detail through the use of Ca^{2+}sensitive fluorescent probes in time lapse imaging experiments. These experiments showed the presence of two distinct peaks in cytosolic Ca^{2+}, corresponding with the pollen tube rupture and egg cell fertilisation respectively (Hamamura *et al.* , 2014). Interestingly though, the fertilisation of the central cell did not result in a Ca^{2+} peak other than the peak associated with pollen tube rupture Hamamura *et al.* (2014).

The orientation of the zygote polarisation then directs the first asymmetric division of the zygote into the 2-cell stage, forming the first cells of the suspensor (larger, micropylar cell) and the embryo proper (EP) (smaller, antipodal cell) (Figure1.3). This asymmetric division is crucial for correct embryo pattern formation, a process characterised by the coordinated spatial differentiation of the suspensor and embryo proper cells. Detailed electron microscopy studies have shown that the precursors for both the suspensor and the EP go through several rounds of cell division, leading into the proembryo stage (Figure 1.2 (top) and Figure 1.3) (Diboll, 1968; Van Lammeren, 1986a). During this stage, pattern formation continues with the elongation of the suspensor, yielding cells that are large and highly vacuolated, while the EP continues dividing into a dense structure comprised of small cells enriched in cytoplasm (Schel *et al.* , 1984).

The final process of phase one, is the progression of the embryo into the transition stage, characterised by the formation of the EP epidermis, known as the protoderm (Figure 1.3). Several genes have been identified which are thought to play a part in the differentiation of the protoderm. The expression of *ZmOCL1-5* (*Outer Cell Layer*) is exclusive to the protoderm, with each gene displaying a different expression pattern in discrete cellular sectors (Ingram *et al.* , 2000). The gene *Lipid transfer protein2* (*Ltp2*) is similar in its expression profile, being restricted exclusively to the abaxial face of the protoderm, albeit at a later stage in embryo development (Sossountzov *et al.* , 1991). These expression zones suggest that the protoderm is not one continuous, uniform cell layer, but is developmentally organised into specialised regions. Furthermore, the *ZmOCL1-5* genes encode putative transcription factors, and are expressed in the outer cell layer immediately before protoderm differentiation, which could indicate that they are involved in a developmental signalling pathway (Ingram *et al.* , 2000, 1999).

Figure 1.2: Seed development stages in maize. Schematic representations of the developmental stages of the overall seed, endosperm and embryo. EC = egg cell, CC = central cell, ESR = embryo surrounding region, BETL = basal endosperm transfer layer, AC = antipodal cells and EP = embryo proper. Drawing not to scale. Adapted from (Hajduch *et al.* , 2005).

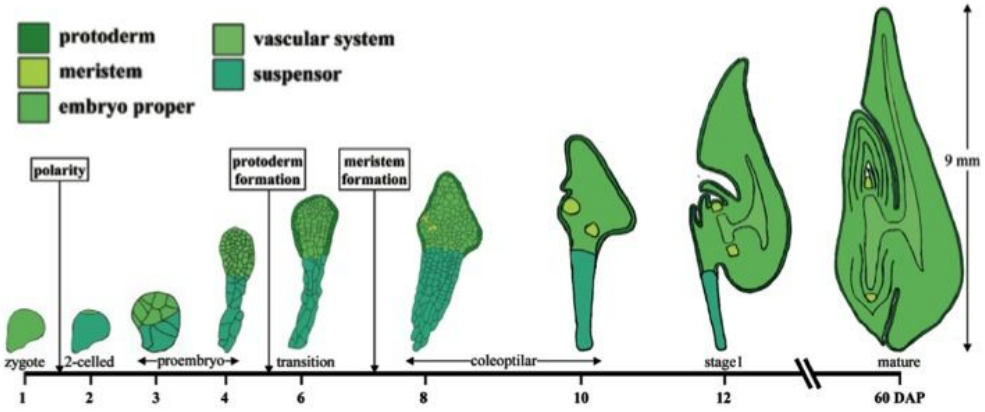

Figure 1.3: Schematic representation of the developmental stages of the maize embryo. Different cell types are indicated by colour scheme and the approximate developmental stage is annotated by the number of days after pollination (DAP). Drawing not to scale. Adapted from Hajduch *et al.* (2005).

The tightly controlled differentiation processes of the transition stage are thought to be regulated in part by exogenous auxins. This was illustrated by local auxin concentrations and the localisation of the auxin transporter *Zm*PIN1a in developing embryos, as determined by a combination of immunohistochemistry and transgenic reporters (Forestan *et al.* , 2010; Chen *et al.* , 2014). Using the auxin response marker DR5, inferred accumulation of auxin was observed on the adaxial, apical side of the developing embryo. This was coupled with detection of the *Zm*PIN1a transporter in a polarised cellular distribution gradually transiting from the site of external auxin accumulation, towards the centre of the embryo proper as it develops (Chen *et al.* , 2014; Dresselhaus & Doughty, 2014).

The second phase of embryogenesis begins with the growth of the embryo and recruitment of nutrient stores while the final developmental processes continue. As the embryo matures, asymmetric outgrowths appear on the adaxial and abaxial sides and the embryo enters the coleoptilar stage (Figure 1.3). This coincides with a distinct expression pattern of PIN proteins in the early shoot/root apical meristems as well as on the

central axis of the early scutellum, indicating a pivotal role for auxins in the development of the late embryo morphology (Forestan *et al.* , 2010; Forestan & Varotto, 2012). The final stages of embryo maturation coordinate the development of the mature scutellum and the embryo axis. This allows the development of the vascular tissues in the scutellum, the coleoptile containing the leaf primordials and the choleorhiza which protects the root apical meristem. In parallel to these final developmental processes, programmed cell death (PCD) initiates in the tissues that are no longer required in the mature embryo. By visualisation of DNA fragmentation, PCD was observed in the suspensor, progressing in a micropylar direction, making way for the developing root apical meristem and surrounding choleorhiza (Giuliani *et al.* , 2002; Hajduch *et al.* , 2005). Once the embryo maturation is complete, the final phase is initiated in which embryo dormancy initiates and dehydration occurs (Hajduch *et al.* , 2005).

1.2.3 The Embryo Suspensor - An umbilical role

The previously mentioned first zygotic division gives rise to two distinct cell types: the embryo proper and the suspensor. Embryo suspensors are a connecting tissue between the embryo proper and the seed coat present in almost all seed bearing plants, a conservation which implies functional importance (Kawashima & Goldberg, 2010). While the general presence of the suspensor is universal, contrastingly, its morphology is remarkably variable amongst different species with regards to dimensions, shape and cell numbers (Raghavan, 2006). Functionally, the suspensor is thought to act primarily as a nutrient and signalling transport tissue, connecting the embryo proper to the nutritive endosperm, similar perhaps to the mammalian umbilical cord. Evidence for this role can be found in the spatial position and cytological description of the suspensor cells. The cells present adaptations that are typical of transfer tissues, such as cell wall ingrowths/outgrowths and abundant plasmodesmata, facilitating efficient

solute transfer (Yeung & Meinke, 1993; Raghavan, 2006; Kawashima & Goldberg, 2010). Furthermore, labelled metabolic tracing experiments have directly shown the directional transit of metabolites from the suspensor to the embryo proper (Yeung, 1980; Nagl, 1990). This was further compounded by evidence that the embryo proper and suspensor form one continuous symplastic compartment as demonstrated by GFP localisation experiments (Stadler *et al.* , 2005).

1.2.4 Endosperm Development - A tale of altruistic sacrifice

Angiosperms have evolved a unique nutritive endosperm which develops after fertilisation, limiting resource investment into potentially non-viable seeds. In the stages immediately following fertilisation, the endosperm development is much more rapid than that of the embryo, giving rise to a tissue that, in monocots, persists until germination. In the dicots, the endosperm is almost completely reabsorbed during cotyledon development. The sole function of the endosperm is to supply the developing embryo with nutrients and so is analogous to the placenta in eutherian mammals (Olsen, 2007).

Three main developmental models have been applied to endosperm development; nuclear, cellular and helobial, reflecting different modes and temporal cues for cellularisation (Floyd & Friedman, 2000). Nuclear is the least common, but is the system adopted by the cereal crops and the model organism *Arabidopsis thaliana*. In the nuclear model, fertilisation of the central cell leads to a series of nuclear divisions without cellularisation, forming a large, free-nuclear structure known as the syncytium. In parallel, the fertilisation of the egg cell initiates a slower but more morphologically structured zygotic development into the mature embryo (Olsen, 2007). In cellular formation, the nuclear division event is immediately followed by cell wall formation, leading the progressive formation of a cellularised endosperm tissue. Helobial development could be considered a hybrid of the two systems mentioned thus far. The initial nuclear division

is followed by cell wall formation, leading on to further nuclear divisions within the new cells. The cellularisation of the endosperm then follows with cell wall deposition along the nuclear division pattern. Central to these models, is that in direct contrast with the gymnosperms, the proliferation of the central cell into the endosperm is actively suppressed until fertilisation by genes in the polycomb group such as *MEDEA/FIS1* and *FIS3/FIE* (Sørensen *et al.* , 2001; Ohad *et al.* , 1999; Kiyosue *et al.* , 1999).

Once the suppression of endosperm development is lifted, the syncytial stage is initiated, lasting for 8-9 rounds of nuclear division before cytokinesis commences. During these rounds of mitosis, the free nuclei migrate in the syncytium to pre-determined locations, known as nuclear cytoplasmic domains (NCDs), in preparation for the formation of specialised tissue sectors. These NCDs are found surrounding the large central vacuole and persist until the onset of cellularisation. Cytokinesis then begins with anticlinal cell wall formation occurring at the central cell periphery between adjacent nuclei, followed by fusion to the nearby central cell wall. This initial row of outer cells later forms the aleurone, a single cell layer defining the outer extremes of the endosperm. Cytokinesis then continues in the absence of nuclear division, sweeping towards the central cell until cellularisation is complete (Olsen *et al.* , 1995; Brown *et al.* , 1996b,a; Nguyen *et al.* , 2001; Olsen, 2007).

1.2.5 Differentiation of the Endosperm Tissues - Specialism defined

Once differentiated, the endosperm of cereals consists of 4 specialised tissue types: The central starchy endosperm, the basal endosperm transfer layer (BETL), the aleurone and the embryo surrounding region (ESR) (Figure 1.2). The starchy endosperm is arguably the most agronomically important feature of cereal grains, containing the highest amount of nutritional content. The cells of the central endosperm are large and specialised for the storage of starch in amyloplasts, contributing to up to 70 % of the

dry seed weight (Olsen, 2007). Some protein is also stored in protein storage bodies within the starchy endosperm cells, contributing further to the nutritional value of this seed tissue (Müntz, 1998; Lopes & Larkins, 1993; Olsen, 2007). Once fully developed, the starchy endosperm then undergoes programmed cell death, driven by the ethylene pathway. This has been shown in the *shrunken2* mutation, which causes premature cell death through the over-production of ethylene (Young *et al.* , 1997; Olsen, 2007). During germination, the starchy endosperm serves as a sacrificial tissue, undergoing rapid enzymatic degradation to provide nutrients for the nascent plant.

1.2.6 Aleurone - The cells of the front-line

As one of the four major cell types in the endosperm, the aleurone is critical to plant survival. The aleurone forms the outer periphery of the endosperm, developing during the initial cytokinesis events in the syncytial endosperm and serves a dual function. During seed development the aleurone functions as a site of storage, while during germination, it functions to rapidly degrade the endosperm for nutrient supply to the developing plant (Young & Gallie, 2000). As a storage tissue, the aleurone accumulates phytic acid, an essential mineral chelator which has been shown in barley to sequester large amounts of phosphates, magnesium, potassium and calcium (Stewart *et al.* , 1988; Olsen, 2007).

Mature aleurone cells are densely packed with endoplasmic reticulum, mitochondria and both lytic and protein storage vacuoles, together with protein and lipid rich globoid bodies and carbohydrate-protein inclusion bodies (Jones, 1969; Jacobsen *et al.* , 1971; Morrison *et al.* , 1975; Kyle & Styles, 1977; Swanson *et al.* , 1998). These features are presumably to support the aleurone's later degradative role, where a rapid supply of energy and enzymes are required. This second role is induced by seed imbibition, initiating a gibberellic acid mediated functional switch. This new role involves the rapid

secretion of hydrolyses such as α-amylase into the starchy endosperm in order to release stored nutrients for use by the developing plant (Bethke *et al.* , 1999; Olsen, 2007).

The development of the aleurone is regulated by positional effects, ensuring that only the outer most cell layer adopts aleurone cell fate. These positional factors appear to be derived from the endosperm itself and are a result of cell extremity rather than from interaction with the exterior nucellus (Olsen, 2007). This has been shown by the formation of aleurone layers in mutant endosperms where the external surface is not in contact with the maternal tissues (Olsen, 2004). Furthermore, endosperm cultures are able to develop aleurone layers in the absence of any other tissue factors (Gruis *et al.* , 2006).

1.2.7 BETL - The nutrient trading-floor

Of the four main tissue types of the endosperm, together, both the aleurone and BETL form a complete epidermis. The BETL, as its name suggests, forms a layer of transfer tissue that interfaces between the maternal and filial tissues (Figure 1.2). The BETL region adjoins a region of compact, dead cells at the terminus of the maternal vascular tissue, known as the placentochalazal region (Kladnik *et al.* , 2004). The outermost cell layer of the endosperm, directly in contact with the placentochalazal region is initially indistinguishable in morphology from the rest of the aleurone. However, following developmental cues, at around 6-8 days after pollination (DAP) this cell layer begins differentiation into the primary transfer cell layer (Thompson *et al.* , 2001). The development of the BETL then continues in a unidirectional mode towards the central endosperm, until about 16 DAP, when it is 3-6 cells deep and fully developed (Thompson *et al.* , 2001; Gao *et al.* , 1998).

The cells of the BETL are named transfer cells (TC), a cell type which functionally resemble the chorionic villi of mammalian placentae, characterised by the existence of

cell wall ingrowths. These ingrowths form a labyrinthine structure, serving to displace the plasma membrane and create an effective transfer tissue with a huge surface area, enhancing the rate of nutrient translocation into the endosperm (Pate & Gunning, 1972; Thompson *et al.* , 2001; Zheng & Wang, 2010; Offler *et al.* , 2003). Further characterisation of the transfer cell identified numerous plasmodesmata in the innermost cells and those at the BETL periphery, thought to be responsible for lateral distribution of the transferred solutes (Davis *et al.* , 1990; Olsen, 2007).

Interestingly, BETL cell identity is not characterised by morphology, but rather by the expression of BETL specific genes (Olsen, 2007). As expected, the expression profile of the BETL tissue is rich in sucrose transporters and cell wall invertases, though none of these genes have been shown to demonstrate transfer cell specificity. Activity of a cell wall invertase has been shown in the characterisation of the *Miniature-1* (*Mn-1*) mutant in maize. *Mn-1* is deficient in a cell wall invertase, *Zm*INCW-2, expressed in the BETL during the large scale metabolite influx occurring at 12-18 DAP (Chourey *et al.* , 2006). This mutation leads to a disruption in nutrient transfer to the endosperm, ultimately resulting in kernels with severely reduced mass (Miller & Chourey, 1992; Cheng *et al.* , 1996).

The aptly named *BETL1-4* genes were the first TC specific genes to be discovered, expressed exclusively in the BETL, even before TC morphology is established. (Hueros *et al.* , 1995, 1999b). These genes encode small cysteine rich proteins (CRPs), a feature shared with several other BETL specific genes such as *ZmMeg1* (Gutierrez-Marcos *et al.* , 2004), *ZmEBE2* (Magnard *et al.* , 2003) and *ZmAE1* (Magnard *et al.* , 2000). The *BETL1-4* gene products have been implicated in defensive roles, owing to their antimicrobial activities and similarity in sequence to the defensin peptides (Olsen, 2007). In fact, *BETL2* was renamed *basal layer anti-fungal peptide 2* (*BAP-2*) on account of its fungicidal properties and its accumulation in a barrier-like formation in the pla-

centochalazal region (Serna *et al.* , 2001).

Given the specificity of these discussed TC genes and the clear spatial definition of the BETL region, a TC specific regulatory pathway was predictably discovered. Two genetic regulatory genes, *Transfer Cell Response Regulator-1* (*ZmTCRR-1*) (Muñiz *et al.* , 2006) and *MYB Related Protein1* (*ZmMRP1*) (Gómez *et al.* , 2002) with expression profiles exclusive to the BETL were identified. Furthermore, (*ZmMRP1*) was characterised as a powerful transcription factor, readily trans-activating the expression of BETL1-2 (Gómez *et al.* , 2002), *ZmMeg1* (Gutierrez-Marcos *et al.* , 2004) and the aforementioned (*ZmTCRR-1*) (Muñiz *et al.* , 2006), demonstrating it as an important regulator of TC identity.

Transfer cell development itself has been shown to be positively regulated in an imprinted fashion by *ZmMeg1* (Costa *et al.* , 2012) and hence by extension, *ZmMRP1.* In the absence of these proteins, TC development is severely reduced, leading to a reduced grain weight coupled with an altered sugar content in the developing seed (Costa *et al.* , 2012). *Zm*MEG1 knockdown was also shown to debilitate the function of the BETL region through the significant reduction of cell wall invertase activity. This led to a reduction in the levels of the hexose sugars, glucose and fructose, normally produced by the inversion of sucrose. Consequently, sucrose levels in the developing seed were found to accumulate far beyond the normal levels in the later stages of development (Costa *et al.* , 2012). This study also linked TC development to sugar signalling, since spatial gradients of glucose have been shown to positively regulate cell proliferation (Wang & Ruan, 2012a).

1.2.8 The Embryo Surrounding Region - To protect and serve

The final distinct tissue of the endosperm, the embryo surrounding region (ESR), is probably the least well understood. The ESR is a transient domain of the central

endosperm juxtaposed to the embryo, present from 3-4 DAP until about 12 DAP, after which it degrades to the benefit of the starchy endosperm and the BETL (Figure 1.2). The first reference to an embryo surrounding region was in 1952 in a research article by Kiesselbach and Walker which investigated the "anatomical and chemical character of certain tissues of the kernel of corn" (Kiesselbach & Walker, 1952). A later study by Schel *et al* captured this ESR tissue in great detail using a combination of light and electron microscopy, describing the cells as being small, densely cytoplasmic and with thin walls. Schel goes further to describe the cells on a sub-cellular level, detailing the abundance of small vacuoles, plasmodesmata, rough endoplasmic reticulum and invaginations in the plasma membrane, though interestingly, in the absence of cell wall ingrowths as seen in the BETL.

The cells of the ESR were assumed to be metabolically and synthetically very active, due to the presence of a large number of mitochondria and vesicles budding from the Golgi body. Furthermore, a polarisation was seen in the organelle organisation, suggesting a directional efflux of metabolites, signals and nutrients towards the developing embryo (Schel *et al.* , 1984). These observations led to the hypothesis that the ESR functions in some sort of nutritive and/or signalling capacity. Significantly, it was determined microscopically that despite the abundant plasmodesmata in the ESR cells, the developing embryo is completely symplastically sealed from the ESR (Lammeren, 1987). This means that the only transit route into the embryo is via the apoplast, effectively creating a passive protective barrier for the embryo and implicating the ESR cells further in secretory activity. However, there are examples in other species of plasmodesmata connecting the symplast of the ESR and embryo at the very early stages of development, though there is, as yet, no evidence of symplastic transport actually occurring (Kozieradzka-Kiszkurno & Płachno, 2011).

In the absence of loss-of-function mutants or transgenics, the definitive function of the

maize ESR is as yet, only hypothetical. However, several lines of evidence have hinted at a range of roles for the ESR through characterisation of ESR specific genes. The first set of ESR genes in maize were suitably named *Embryo Surrounding Region 1-3* (*ESR1-3*), three highly homologous genes showing an ESR specific expression profile demonstrated by both mRNA hybridisation and promoter-GUS reporter fusions (Opsahl-Ferstad *et al.* , 1997; Bonello *et al.* , 2000). The gene products of the *ESRs* were later also shown to localise to the ESR and additionally, to be secreted from the cell to interact with the cell wall (Bonello *et al.* , 2002). Though the exact function of these genes remains unknown, the high degree of homology and the identical expression patterns strongly suggest that *ESR1-3* are functionally redundant isoforms, conservation of which indicates functional importance (Bonello *et al.* , 2000).

Studies investigating the gene expression profiles associated with maize androgenesis identified two genes *ZmAE1* and *ZmAE3* that are ESR/BETL and ESR specific respectively (Magnard *et al.* , 2000). These genes encode small CRPs with homology to the previously discussed anti-fungal protein BAP-2 (Magnard *et al.* , 2000; Sevilla-Lecoq *et al.* , 2003; Serna *et al.* , 2001). Furthermore, another defensin-like CRP, *Zm*ESR6, was identified showing a characteristic ESR-specific expression and strong anti-microbial activity *in vitro* (Balandín *et al.* , 2005). These data suggest a defensive role for the ESR, protecting the vulnerable developing embryo from pathogenic attack, though *Zm*ESR6 may have an unidentified signalling role and the antimicrobial function could be a side effect of the defensin-like protein structure.

Two more genes showing ESR expression characteristics are *ZmEBE1-2*, though these are not entirely ESR specific, also displaying expression in the BETL (Magnard *et al.* , 2003). Interestingly though, *ZmEBE1-2* showed a novel expression pattern in maize, being expressed both in the central cell before fertilisation and the BETL and ESR in the developing fertilised seed. Again, due to a lack of knock-out studies, the function

of *ZmEBE1-2* remain hypothetical, perhaps involved in BETL/ESR development or eliciting a signalling role.

Further evidence for a signalling function for the ESR came with the discovery and characterisation of the *ZmINVINH1* gene which encodes one of the three invertase inhibitors identified in maize. Invertases function at critical sugar translocation points such as the BETL to hydrolyse the disaccharide sucrose into the hexose sugars; glucose and fructose. As demonstrated by the *Mn1* mutant, this function has a direct effect on grain yield due to the synergistic actions of glucose as a signalling molecule, and plant signalling hormones such as auxin, promoting cell division and inhibiting differentiation (Wang & Ruan, 2012a, 2013; Ishimaru *et al.* , 2013; Wobus & Weber, 1999; Weschke *et al.* , 2003). This has led to the "invertase control hypothesis" in which aspects of seed development are influenced by the temporal synchronisation between developmental cues and the environmentally induced availability of sucrose as a metabolite (Bate *et al.* , 2004; Wobus & Weber, 1999; Costa *et al.* , 2012).

The maize invertase inhibitors are CRPs identified by homology with similar CRPs found in rice, wheat, tobacco and *Arabidopsis* which exhibit characteristic activity attenuation in the presence of sucrose (Bate *et al.* , 2004; Sander *et al.* , 1996). This feature creates a self-regulating feedback system in which sucrose hydrolysis is promoted when the metabolic status of the plant provides an abundance of the invertase substrate, sucrose. The subsequent increase in the product glucose then promotes cell division, effectively increasing carbon sink-strength in environmentally favourable situations. In contrast to the other two maize inhibitors, expression of *ZmINVINH1* is restricted specifically to the apoplast of the ESR, indicating a role in the regulation of endosperm/embryo development rates by shielding the embryo from the high glucose concentrations found in the endosperm (Bate *et al.* , 2004). Invertase inhibitor activity was demonstrated by the *in vitro* inhibition of both soluble (vacuolar) and insoluble

(cell wall) maize invertases, with a preferential effect on the insoluble invertase. High level expression of the soluble invertase IVR2 occurs in the ESR, an expected behaviour, given the rapid expansion of the early ESR, however, activity of the cell wall invertases INCW1-4 has not been detected (Chourey *et al.* , 2006; Kim *et al.* , 2000). Therefore, the presence of the inhibitor *ZmINVINH1* suggests either an as yet unidentified ESR invertase, or a precautionary role for *ZmINVINH1*, safeguarding against ectopic invertase activity in the ESR (Bate *et al.* , 2004).

In other species, there also exists regions of apparently specialised cells that surround the developing embryo during the early stages. In wheat, a cytologically similar tissue has been observed which develops rapidly in close proximity to the developing embryo, referred to as the "modified endosperm". Similarly, the analysis of developing barley caryopses has identified a region of distinct cells situated adjacent to the developing embryo, showing many similar features to those of the discussed maize ESR (Brown *et al.* , 1994; Engell, 1989). While these tissues show similar characteristics, since no conclusive functional studies have been carried out, the comparisons are merely speculative at this point. In *Arabidopsis*, the homologous tissue type is described as the micropylar endosperm (MCE), a region of endosperm cells that are similar in cytoplasmic density and envelop the nascent embryo (Brown *et al.* , 2003). Furthermore, genetic zoning can be observed in the MCE in a similar fashion to that which is observed in maize, with preferential or exclusive expression of certain genes. Experiments using GUS promoter fusions localised the expression of the *Miniseed3* promoter, a WRKY family gene that is required for normal seed development, to the *Arabidopsis* MCE (Luo *et al.* , 2005). Furthermore the promoter of the sucrose transporter *AtSuc5*, was fused to a *Gfp* reporter and upon analysis, was shown to have a spatial expression profile restricted to the MCE in the early stages of development (Baud *et al.* , 2005).

Considering both that the ESR completely envelops the nascent embryo and that the

embryo is unable to develop without the presence of the surrounding tissues, the endosperm factors required by the embryo must either originate from, or transit through, the ESR (Leduc *et al.* , 1996; Mòl *et al.* , 1993; Olsen, 2007). An example of a gene set expressed in the ESR of another plant is the *Embryo Surrounding Factor1* (*ESF1*) gene family in *Arabidopsis* (Costa *et al.* , 2014). These recently characterised genes encode small CRPs with a high degree of structural homology to the *Zm*MEG1 proteins in the maize BETL. Interestingly, the *ESF1* family is expressed both in the unfertilised central cell and in the micropylar endosperm which surrounds the early developing embryo and is thought to be functionally equivalent to the ESR in maize. Significantly, the ESF1 proteins were found to be crucial in the normal pattern formation of the embryo, positively regulating the elongation phase of suspensor development. Through genetic outcrossing experiments, ESF1s were found to act in the same signalling pathway as the previously characterised kinase genes *YODA* (*YDA*) and *Short Suspensor* (*SSP*), both of which are critical for correct asymmetric zygote division and hence establishment of normal embryonic patterning (Costa *et al.* , 2014; Bayer *et al.* , 2009; Lukowitz *et al.* , 2004).

1.3 Small Cysteine-Rich Proteins in Plants - Jack of all trades

Many of the developmental processes discussed thus far have involved the expression or action of CRPs, a class of proteins which until relatively recently has been significantly underestimated. Computational work on the genomes of *Medicago truncatula*, *Lotus japonicus*, soybean (*Glycine max* and *Glycine soja*), rice (*Oryza sativa*) and *Arabidopsis* identified a gene family coding for proteins with the following characteristics; "(i) an N-terminal signal peptide, (ii) a small divergent charged or polar mature peptide with conserved cysteines, (iii) a similar intron/exon structure, (iv) spatial clustering in the genomes studied, and (v) overrepresentation in expressed sequences from reproductive

structures of specific taxa" (Silverstein *et al.* , 2007). This analysis yielded over 13 000 plant genes forming the CRP family, representing up to 3% of the gene compendium in a given model organism (Silverstein *et al.* , 2007). Classically, plant signalling has been attributed to hormones such as auxins and cytokinins which elicit responses at long range. Unlike in the mammalian systems, few examples of peptide/receptor based signalling pairs have been characterised in plants (De Smet *et al.* , 2009). However, this emerging field has identified hundreds of candidate receptor like kinases (RLKs) and putative secreted peptides which could be acting as signalling ligands for short range cell communication (Butenko *et al.* , 2009).

Given the enormity of this CRP family, it is unsurprising that a wide range of functions have been accredited to its members (for a selected summary, see Table 1.1). CRPs have been implicated in defence, bacterial symbiosis, development, pollen recognition, pollen guidance and seed function across a wide range of plant species (Marshall *et al.* , 2011). The plant defensin family makes up a significant portion of the defensive peptides produced in plants, characterised as small, secreted CRPs that undergo post-translational processing into short, antimicrobial peptides with conserved γ-core and CSαβ-motif regions (Hammami *et al.* , 2009; Zhu *et al.* , 2005). More than 300 defensins with features that resemble anti-microbial peptides have been identified in *Arabidopsis* alone (Silverstein *et al.* , 2007). As previously discussed, the maize CRPs *Zm*ESR6 and *Zm*BAP2 are expressed in the ESR and BETL respectively and also both demonstrate *in vitro* antimicrobial activity (Balandín *et al.* , 2005; Serna *et al.* , 2001). However, there has been no direct evidence showing antimicrobial activity *in planta*, leaving the possibility that this property could be a coincidental feature of this class of proteins.

Another role for CRPs can be found in the symbiotic world of root nodulation. Plants in the legume family such as beans and peas exhibit a mutually beneficial relationship with nitrogen-fixing bacteria in the *Rhizobiaccae* family, which invade the root network

of the plant. Signal molecules called Nod factors are released by the bacteria and recognised by the host plant, which initiates a developmental process to envelop the bacteria, forming root nodules. Several nodule-specific CRPs have been identified, including ENOD3 and ENOD14 which appear to play an important role in nodulation, owing to their exclusive expression during the early developmental stages (Scheres *et al.* , 1990b). In *Medicago*, genomic based studies have identified more than 300 nodule-specific CRPs (NCRs) which demonstrate highly variable sequences that contain four conserved cysteine residues, and an overall structure that his homologous to NCRs identified in other legumes (Mergaert *et al.* , 2003; Graham *et al.* , 2004; Scheres *et al.* , 1990a). Interestingly, it is becoming more apparent that these CRPs probably function in a signalling role, perhaps acting as receptor ligands, an hypothesis supported by the presence of a nodule-specific secretory pathway containing the signal peptidase *dnf1-1* which seems to regulate terminal bacterial differentiation in nodulation (Van de Velde *et al.* , 2010; Wang *et al.* , 2010).

Another key biological role demonstrated by CRPs is in the area of plant development. The *Rapid Alkalinization Factor* (*RALF*) gene family produces secreted CRPs that can be found in the roots of tobacco, sugarcane and *Arabidopsis*, appearing to play a key role in root development, ensuring normal growth and nutrient uptake capacity (Pearce *et al.* , 2001; Mingossi *et al.* , 2010; Olsen *et al.* , 2002). This role has recently been characterised further, demonstrating that in *Arabidopsis*, the RALF protein activates the FERRONIA cell surface receptor (Haruta *et al.* , 2014). This interaction leads to the phosphorylation of H^+-adenosine triphosphatase 2, resulting in reduced proton transport and an associated increase in extracellular alkalinity (Haruta *et al.* , 2014). At the opposite extremes of the plant, CRPs can be found with a role in development of stomatal patterning in the leaves, participating in a receptor-mediated signalling network (Nadeau & Sack, 2002; Shpak *et al.* , 2005). The cysteine-rich Epidermal Patterning

Factors in *Arabidopsis* (*At*EPF1-3) have critical functions in the regulation of the frequency and distribution of stomata through spatial coordination of the primordial cells that develop into the stomatal components (Hara *et al.* , 2009; Kondo *et al.* , 2010; Sugano *et al.* , 2010). These EPFs act synergistically with the receptors ERECTA (ER), ERECTA-LIKE1 (ERL1), ERECTA-LIKE2 and TOO MANY MOUTHS (TMM) in a signalling pathway that controls stomatal patterning (Shpak *et al.* , 2005). The EPF-related *Arabidopsis* proteins CHALLAH and STOMAGEN also function in the same regulatory system, though interestingly, CHALLAH is another negative regulator while STOMAGEN is a promoter of stomatal development, indicating a finely controlled signalling network (Abrash & Bergmann, 2010; Sugano *et al.* , 2010; Kondo *et al.* , 2010; Jewaria *et al.* , 2013; Richardson & Torii, 2013).

1.3.1 Reproductive roles for CRPs - From speed-dating to wedlock

CRPs have been discovered to have a variety of functions that are essential in the various stages of double fertilisation. In the environment, there are many opportunities for pollen from other species to come into contact with the female stigma, so mechanisms exist to prevent fertilisation in these circumstances. However, outcrossing is essential for evolving genetic diversity, and therefore plants have developed a mechanism for preventing self fertilisation known as self-incompatibility. In *Brassica* a small CRP called SCR/SP11 is expressed in the anthers and pollen coat and functions as a signalling peptide in self-recognition through interaction with two female receptor kinases present on the stigma MLPK and SRK, the presence of which prevents self-pollination (Schopfer *et al.* , 1999; Suzuki *et al.* , 1999; Stein *et al.* , 1991; Murase *et al.* , 2004; Takayama *et al.* , 2001). A similar mechanism exists in the Poppy, but the small CRP, PrsS, is expressed in the stigma of the female plant and interacts with PrpS, a transmembrane protein expressed in the pollen (Wheeler *et al.* , 2009).

Assuming that successful pollen recognition occurs at the stigma, the next reproductive process involves the growth and guidance of the pollen tube, navigating its way through the female ovary to deliver the male gametes in order for fertilisation to succeed. The small CRPs LAT52 and STIG1 have been demonstrated to positively regulate pollen tube initiation and growth in the stigma and style respectively through interaction with pollen specific Leucine Rich Repeat (LRR) receptor kinases (Muschietti *et al.* , 1994; Tang *et al.* , 2004). Through laser ablation studies, the directional cues for the guidance of the pollen tubes have been shown to originate from the synergids (Higashiyama *et al.* , 2001). At least one synergid was required to attract the pollen tube, though after fertilisation, despite the persistence of the second synergid, pollen tube attraction immediately halts. This is an indication of a mechanism which safeguards against polyspermy and could be a downstream signalling consequence of the previously discussed $[Ca^{2+}]$ pulse in the egg cell upon fertilisation (Hamamura *et al.* , 2014). More recently, work in *Torenia* has identified LURE1/2, two CRPs expressed in the synergids that positively attract the growing pollen tube in a species specific manner (Okuda *et al.* , 2009). Interestingly, the protein Egg Apparatus 1 (*Zm*EA1) expressed in the synergids and egg cell seems to fulfil an equivalent attractant function in maize, but is not a member of the CRP family (Uebler *et al.* , 2013; Márton *et al.* , 2005).

On arrival of the pollen tube at the egg cell, expulsion of the sperm cells occurs through rupture of the pollen tube itself. In *Arabidopsis,* the pollen and synergid specific receptors, ANUXUR1/2 and FERONIA respectively, have been demonstrated to be essential for pollen tube discharge but their ligands have not yet been identified (Boisson-Dernier *et al.* , 2009; Escobar-Restrepo *et al.* , 2007; Miyazaki *et al.* , 2009). However, as previously mentioned, FERONIA has been shown to specifically interact with RALF in the roots (Haruta *et al.* , 2014), so it is possible that a similar interaction is occurring during the processes of fertilisation, though further work is required to demonstrate this.

A CRP candidate *Zm*ES4, has been identified in maize which is able to rapidly induce species-specific pollen tube rupture *in vitro* and demonstrates an associated function in the opening of the K^+ion channel KZM1 (Amien *et al.* , 2010). Furthermore, a second CRP expressed in both the male and female maize gametophytes encodes a pectin methylesterase inhibitor (*Zm*PMEI1) has been shown to induce pollen tube rupture *in vitro* through the destabilisation of the cell wall (Woriedh *et al.* , 2013). Interestingly though, ZmPMEI1 appears to be broad spectrum, inhibiting PMEs from unrelated species. Additionally, there are several CRPs involved in seed function that have already been discussed, such as *Zm*MEG1, *At*ESF1, *Zm*ESR6 and *Zm*INVINH1, demonstrating a varied and crucial role for this fascinating class of proteins.

Table 1.1: Plant CRPs and their putative functions in different plant species. Adapted from (Marshall *et al.* , 2011).

Name	Plant	Expression	Role	Function	Reference
EPF1	*Arabidopsis*	Leaf epidermis	Development	Stomatal positioning	(Hara *et al.* , 2007)
EPF2	*Arabidopsis*	Leaf epidermis	Development	Negative regulator of stomatal density	(Hara *et al.* , 2009; Hunt & Gray, 2009)
STOMAGEN	*Arabidopsis*	Leaf mesophyll	Development	Positive regulator of stomatal density	(Sugano *et al.* , 2010; Kondo *et al.* , 2010)
CHALLAH	*Arabidopsis*	Leaf epidermis	Development	Negative regulator of stomatal density	(Abrash & Bergmann, 2010)
RALF	*Arabidopsis*	Roots	Development	Regulator of extraxellular pH	(Haruta *et al.* , 2014)
ESF1	*Arabidopsis*	Endosperm	Seed function	Positive regulator of embryo patterning	(Costa *et al.* , 2014)
LAT52	Tomato	Pollen	Reproduction	Pollen tube germination	(Muschietti *et al.* , 1994)
STIG1	Tomato	Pistil	Reproduction	Positive regulator of pollen tube growth	(Tang *et al.* , 2004)
SCA/LTP5	Lily/ *Arabidopsis*	Stigma/style	Reproduction	Pollen adhesion	(Mollet *et al.* , 2000; Chae *et al.* , 2009)
PrsS	Poppy	Stigma	Reproduction	Female SI determinant	(Wheeler *et al.* , 2009)
SCR/SP11	*Brassica*	Pollen	Reproduction	Male SI determinant	(Schopfer *et al.* , 1999)
LUREs	*Torenia*	Synergid	Reproduction	Pollen tube attractant	(Okuda *et al.* , 2009)
ES4	Maize	Synergid	Reproduction	Pollen tube rupture	(Amien *et al.* , 2010)
PMEI1	Maize	Embryo sac & pollen	Reproduction	Pollen tube rupture	(Woriedh *et al.* , 2013)
EC1	Maize	Egg cell	Reproduction	Sperm cell activation	(Sprunck *et al.* , 2012)
BAPs	Maize	Endosperm BETL	Seed function	Pathogen defence	(Serna *et al.* , 2001)
AE1	Maize	Endosperm ESR/BETL	Seed function	Embryo/endosperm crosstalk?	(Magnard *et al.* , 2000)
AE3	Maize	Endosperm ESR	Seed function	Embryo/endosperm crosstalk?	(Magnard *et al.* , 2000)
MEG1	Maize	Endosperm BETL	Seed function	Positive regulator of TC development	(Costa *et al.* , 2012)
ESR6	Maize	Endosperm ESR	Seed function	Pathogen defence	(Balandín *et al.* , 2005)
INVINH1	Maize	Endosperm ESR	Seed function	Invertase inhibitor	(Bate *et al.* , 2004)
BETL1,2,4	Maize	Endosperm BETL	Seed function	Pathogen defence?	(Hueros *et al.* , 1995)

1.3.2 CRP structure - Constrained disorder

The primary sequence configuration of the small CRPs is broadly conserved: Short sequence lengths (<160 AA) are initiated by N-terminal secretory signals, followed by variable lead sequences and C-terminal cysteine-rich domains (CRDs). The CRDs are composed of disulphide bond forming cysteine pairs, interspaced by hyper-variable stretches of amino acids, a structural format conserved in a variety of proteins, species and indeed, kingdoms (Marshall *et al.* , 2011; Beets *et al.* , 2012). Given the enormity and importance of this protein class as previously discussed, there are surprisingly few examples with a fully solved 3D structure. This is probably due to the difficulties encountered in efficient expression of active forms of CRPs owing to the complex folding pathways required (Nguyen *et al.* , 2011). However, some plant CRPs have been structurally studied including: The radish anti-fungal protein *Rs*AFP1 (Fant *et al.* , 1998), the defensin γ-Thionins from barley and wheat (Beets *et al.* , 2012), *Arabidopsis* STOMAGEN (Ohki *et al.* , 2011), *Brassica* SP11 (Mishima *et al.* , 2003); and most recently, *Arabidopsis* ESF1.3 (Costa *et al.* , 2014). While there are significant variations in the sequences for these proteins, when the 3D structure of the CRPs is considered the disulphide bonds construct a highly stable scaffold enabling the inter-cysteine sequences to form solvent accessible loops. These loops typically contain flexible, disordered regions comprised of highly variable amino acid sequences, suggesting a great degree of plasticity in the CRP class.

The cysteine-rich core is essential for supporting the correct conformation and hence activity of the CRPs, demonstrating that despite their small size, their 3D organisation is crucial (Ohki *et al.* , 2011). The disordered loop regions formed by the inter-cysteine sequences have been shown to contain the essential residues for activity in STOMAGEN (Ohki *et al.* , 2011), suggesting that this scaffold formation may have evolved to support a variety of loop lengths and sequences, the variability of which may be crucial in

conferring the wide range of functions observed. Additionally, the above mentioned CRP structures all contain a mixture of α-helix and β-sheet secondary structure, a conservation which suggests functionality although no experimental evidence has been obtained demonstrating the importance of these secondary structural features.

1.4 Project Aims

The major aim of this project was to generate further understanding of small cysteine-rich plant proteins and their involvement in reproductive development of cereal crops. This work has been carried out in conjunction with Biogemma, in an attempt to glean knowledge that could lead to the generation of higher nutritionally yielding crops. Biogemma is a French biotechnology company and the industrial CASE sponsor for this project.

Specific aims include:

1. Identification and characterisation of a novel CRP gene (ZmMlp1) expressed in maize seeds through studying its promoter activity, spatio-temporal gene expression profiling, protein localisation and biochemical characterisation, placing this information into the context of seed development (Chapter 3).

2. Functional investigation of *ZmMlp1* through the phenotypic analysis of RNAi mediated transcriptional knock-down transgenics, whole-transcriptome shotgun sequencing and genetic complementation studies and to discuss the way in which this data fits into the wider field (Chapter 4).

3. Gain understanding of the key structural features of *Zm*MLP1 through NMR structural determination, cross-species structure-function experiments and protein modelling to further understand the molecular mechanisms involved in seed development and how CRPs play a major role (Chapter 5).

2 Materials and Methods

2.1 General Plant Materials and Methods

2.1.1 *Zea mays* plant growth

Plants were grown in maize mix compost (1 part Sphagnum moss peat, 1 part John Innes No.3, 1 part course grade horticultural sand and 3 g/L Osmocote slow release fertiliser). Plants grown in glasshouse conditions, 16 h day, 8 h darkness photoperiod, 28 °C. Seed germination carried out in 9 cm pots, transferred to 23 cm pots after 3-4 weeks of growth and grown to maturity.

2.1.2 *Nicotiana benthamiana* plant growth

Seeds of *N.benthamiana* were sewn onto pre-wet Levingtons F1 compost (three parts soil, one part vermiculite), covered with a petri dish and germinated in a growth chamber (Conviron) (16 h day, 8 h night photoperiod at 22 °C, light intensity 250 µmol/sec/m^2, humidity 70 %, CO_2 350 ppm) for 7-10 days. Plants were then separated and transferred into individual 9 cm pots for growth to maturity under glasshouse conditions (16 h day, 8 h night photoperiod at 22 °C).

2.1.3 *Arabidopsis thaliana* plant growth

Seeds of *A.thaliana* were vernalised by suspension in 0.1 % sterile agarose and incubation at 4 °C for 2-3 days in darkness. Seeds were then pipetted onto pre-wet Levingtons F1 compost (three parts soil, one part vermiculite), and germinated in a growth chamber (Conviron) (16 h day, 8 h night photoperiod at 22 °C, light intensity 250 µmol/sec/m^2, humidity 70 %, CO_2 350 ppm) for 7-10 days. 2-5 plants were then selected and transferred into a 9 cm pot for growth in the growth chamber until approaching flowering stage. Plants were then transferred to glasshouse conditions (16 h day, 8 h night photoperiod at 22 °C) for growth to maturity.

2.1.4 Transgenic maize plantlet growth

Transgenic plantlets generated by Biogemma were couriered to The University of Warwick in individual culture tubes in growth media. Plantlets were then removed from tubes, the roots washed in water to remove growth media and transferred into 9 cm growth pots in maize mix soil (see 2.1.1). Plantlets were covered in plastic bags to maintain a humid environment for 4-5 days before uncovering and growth to maturity (see 2.1.1).

2.1.5 Maize controlled pollination

Maize crosses were performed manually, individually selecting and pollinating plants. Plants approaching flowering stage were monitored for the emergence of paternal tassels and maternal ear shoots. Nascent ear shoots were covered with a paper shoot bag to prevent environmental pollination and examined daily for silk protrusion. 24 h preceding pollination, silks were cut back to 1 cm in length and re-covered. Existing pollen was removed from male pollinator plants at the end of the day prior to pollination. Pollen was subsequently collected by gentle tassel agitation into a pollen bag and applied directly to the new silk growth on the maternal plant before re-covering. Plants were grown for approximately 4 weeks, until seed maturity was achieved, before cessation of watering, drying of plants and ear harvesting. Seeds were removed from ears and stored at 5 °C and 20 % humidity.

2.1.6 *Arabidopsis* controlled pollination

Immature flower buds of the maternal parent plant were emasculated by removal of the nascent anthers with fine forceps. Pollination was then performed the following day by

manually removing stamens from the paternal plant and transferring pollen onto the female stigma by direct contact.

2.1.7 Generation of pro*Mlp1*:GUS transgenic maize lines

The pro*Mlp1* (1182bp) sequence was sent to Biogemma and synthesised (Genscript USA) flanked by *Kpn*I and *Xba*I cloning sites. This fragment was then cloned in front of a GATEWAY (Invitrogen) aatR1-CmR- CcdB aatR2 Arabidopsis Sac66 polyadenylation sequence. The resulting pro*Mlp1* – Gateway – terminator cassette was then cloned as a *Xho*I / *Pme*I fragment into the binary vector pBIOS 879 (Ds Actin Bar CsVMV FAD2 GFP HvMod) linearised with *Xho*I / *Pme*I. A GATEWAY LR reaction was then used to introduce GUS behind the MLP1 promoter, forming the plasmid pBIOS1865 (see 7.2 figure 7.6). These vectors, based on the plant binary vector pSB12 (Komari *et al.* , 1996), also contain a selectable marker and GFP genes under the constitutive promoter pCsVMV (Verdaguer *et al.* , 1998) in order to visually follow the transgene. The resulting plasmids were transferred into *Agrobacterium tumefaciens* strain LBA4404 (pSB1) as described by Komari *et al.* (Komari *et al.* , 1996). The maize cultivar A188 was transformed with these strains essentially as described by Ishida *et al.* (Ishida *et al.* , 1996). Plant selection was performed using glufosinate. Single copy insertion events were selected via QPCR analysis for the selectable marker gene.

2.1.8 Generation of MLP1-GFP transgenic maize lines

The MLP1:GFP fused sequence was sent to Biogemma and a synthetic gene created (Genscript USA) flanked by *Kpn*1 and *Xba*I cloning sites. The synthetic fragment was then cloned behind the pro*Mlp1* promoter in a binary vector derived from pSB11 resulting in the plasmid pBIOS1865 (see 7.2 Figure 7.7). These vectors, based on the

plant binary vector pSB11 (Komari *et al.* , 1996), also contain a selectable marker and GFP genes (Verdaguer *et al.* , 1998) in order to visually follow the transgene. The resulting plasmid was transformed into *Agrobacterium tumefaciens* strain LBA4404 (pSB1) as described by Komari *et al.* (Komari *et al.* , 1996). The maize cultivar A188 was transformed with these strains essentially as described by Ishida *et al.* (Ishida *et al.* , 1996). Plant selection was performed using glufosinate. Single copy insertion events were selected via QPCR analysis for the selectable marker gene.

2.1.9 Generation of maize pro*Mlp1*:RNAi transgenic lines

A synthetic fragment (Genscript USA) consisting of an inverted repeat of 296 bp of *ZmMlp1* separated by the potato *Stl1* intron was cloned behind the pro*Mlp1* promoter in a binary vector derived from pSB11 resulting in the plasmid pBIOS 2755 (see 7.2 Figure 7.9). These vectors, based on the plant binary vector pSB11 (Komari *et al.* , 1996), also contain a selectable marker and GFP genes (Verdaguer *et al.* , 1998) in order to visually follow the transgene. The resulting plasmid was transformed into *Agrobacterium tumefaciens* strain LBA4404 (pSB1) as described by Komari *et al.* (Komari *et al.* , 1996). The maize cultivar A188 was transformed with these strains essentially as described by Ishida *et al.* (Ishida *et al.* , 1996). Plant selection was performed using glufosinate. Single copy insertion events were selected via qPCR analysis for the selectable marker gene.

2.1.10 Transformation of *Agrobacterium* by electroporation

Agrobacterium (GV3101:pMP90 with Gentamicin) electrocompetent cells were transformed with the relevant construct in a binary plant expression vector (Koncz & Schell, 1986) by electroporation. 30 µl of GV3101 cells were incubated on ice for 10 min with

0.5 µl of the appropriate plasmid. The solution was pipetted into an electroporation cuvette (Bio- Rad) and placed into the electroporator (Bio-Rad Micropulser) before pulsing (Capacitance 25 µF; Voltage 2.4 kV; Resistance 200 Ω; Pulse length 5 msec). Immediately after the pulse, 1 mL of room temperature LB low salt media was added to the cells. Transformations were left to recover at 28 ^{o}C for 3 h before pipetting onto LB low salt agar plates containing appropriate antibiotic selection.

2.1.11 *Agrobacterium tumefaciens* mediated transient expression in Tobacco leaves

A 5 mL culture of transformed *Agrobacterium tumefaciens* (see 2.1.10) was grown from a single colony at 28 ^{o}C, 220 rpm, O/N in LB low salt (10 g L^{-1}tryptone, 5 g L^{-1} yeast extract, 0.5 g L^{-1}NaCl) with gentamicin and vector antibiotic. Two millilitres of culture was centrifuged at at 16000 g for 45 sec, supernatant discarded and pellet resuspended in 1 mL infiltration buffer (55 mM MES, 28 mM glucose, 2 mM Na_3PO_4, pH 5.6). Sample centrifuged at 16000 g for 45 sec, supernatant discarded and pellet resuspended in 1 mL infiltration buffer with 100 µM acetosyringone (4'-Hydroxy-3',5'-dimethoxyacetophenone). Sample centrifuged at 16000 g for 45 sec, supernatant discarded and pellet resuspended in 1 mL infiltration buffer with 100 µM acetosyringone. Sample diluted in infiltration buffer with 100 µM acetosyringone, OD_{600} measured and adjusted to 0.2. In the case of dual transformation by co-infiltration, the two preparations of cells at $OD_{600}$of 0.2 were mixed in 1:1 ratio at this point. The lower leaf epidermis of 6 week old *Nicotiana benthamiana* plants (see 2.1.2) were infiltrated with the solution using a 1 mL syringe, applying gentle pressure above the syringe on the upper leaf surface. The plant was incubated at 20 ^{o}C for 2 days before expression level checked by confocal microscopy of GFP positive control infiltration (see 2.5.5).

2.1.12 Transformation of *Arabidopsis thaliana*

A.thaliana plants were transformed by the floral dip method (Clough & Bent, 1998). A 10 mL culture of transformed *Agrobacterium tumefaciens* (see 2.1.10) was grown from a single colony at 28 ^{o}C, 220 rpm, O/N in LB low salt (10 g L^{-1}tryptone, 5 g L^{-1} yeast extract, 0.5 g L^{-1}NaCl) with gentamicin and vector antibiotic. 5 mL of starter culture was then used to inoculate 500 mL LB low salt with gentamicin and vector antibiotic in a 2 L culture flask and grown at 28 ^{o}C, 220 rpm, O/N. The culture was centrifuged at 5000 rpm and the pellet resuspended in 1 L 5 % sucrose, 0.02 % Silwet L-77 (Lehle Seeds). Closed *Arabidopsis* inflorescences and buds were dipped into the solution for 30 seconds. Plants were incubated on their sides for 24 h in trays covered by clear polythene under glasshouse conditions, before uncovering and growth under normal conditions for 7 days. Plants were then placed into plastic bags and grown to maturity before seed collection.

2.1.13 BASTA selection of transgenic *Arabidopsis thaliana*

Segregating transgenic seeds were selected for positive transformants using the BASTA resistance (BAR) selection marker by germination of seeds on soil pre-soaked in a working solution of BASTA herbicide (150 mg L^{-1}ammonium glufosinate). Seed trays were watered twice weekly with BASTA for 2 weeks before viable seedlings were transferred to individual pots and grown to maturity without further BASTA treatment.

2.1.14 *Arabidopsis thaliana* seed sterilisation

Approximately 200 seeds were placed inside a 1.5 mL tube and 500 µL 7.5 % bleach was added, followed by agitation for 10 mins. Seeds were then washed with 1 mL sterile

dH_2O five times under sterile conditions. Seeds were then suspended in 0.1 % sterile agarose.

2.1.15 Transgene insertion quantification

Between 50 and 100 T_1seeds from BASTA resistant plants were sterilised (see 2.1.14) and individually plated onto MS agar, sucrose, glufosinate media (4.4 g L^{-1}Murashige and Skoog (MS), 3 % (w/v) sucrose, 0.8 % (w/v) phyto agar, 20 µg mL^{-1}glufosinate ammonium (PESTANAL), pH 5.7). Seeds were then vernalised at 4 ^{o}C for 2-4 days in the dark, before transferring to controlled environment chambers for germination (see 2.1.3). After 1-2 weeks, glufosinate resistant seedlings were counted against the total number that germinated. The ratio of resistant:susceptible plants was then calculated to indicate the number of chromosomal insertions and the data analysed by Chi squared test.

2.2 General Molecular Materials and Methods

2.2.1 RNA extraction from maize kernels

100 mg of ground, frozen maize kernels was added to 200 µL of RNA extraction buffer (50 mM Tris-HCl pH8.0, 150 mM LiCl, 5 mM EDTA pH 8.0, 1% SDS). 200 µL of 1:1 phenol chloroform was added, sample shaken and placed on ice for 5 mins, mixing occasionally. Solution then transferred to a pre-spun Phase Lock tube (5 Prime) and centrifuged for 10 mins at 10,000 g, 4 °C. 200 µL phenol-chloroform was added, shaken well and centrifuged for 10 mins at 10,000 g, 4 °C. 200 µL chloroform was then added, shaken well and incubated on ice for 5 mins, mixing occasionally. Samples were then centrifuged for 10 mins at 10,000 g, 4 °C and the aqueous (upper) phase transferred to a new Phase Lock tube. 1 mL TRIzol reagent (Invitrogen) then added, sample shaken

for 15 secs and incubated for 5 mins at RT. 200 µL of chloroform was added, the sample shaken, incubated for 2-3 mins at RT and centrifuged for 10 mins at 10,000 g, 4 °C. 700 µL of the aqueous phase was transferred to a new tube and the RNA was precipitated using 700 µL isopropanol overnight on ice. RNA was pelleted by centrifugation for 10 mins at 10,000 g, 4 °C. RNA pellet was then washed using 1 mL of 70 % EtOH before air drying and dissolving in DEPC treated dH_2O. RNA was then quantified by UV-visible spectrophotometry (NanoDrop).

2.2.2 RNA extraction by hot-phenol method

100 mg of frozen maize endosperm and five 0.2 mm glass beads were added to 500 µl of well mixed hot phenol extraction buffer (0.1 M Tris-HCl pH 8.0, 0.1 M LiCl, 10 mM EDTA pH 8.0, 1 % SDS, mixed in a 1:1 ratio with phenol at 80 °C) and shaken for 30-40 secs in a Ivoclar Vivodent Silamat S6 amalgamator. 250 µl of chloroform:IAA (24:1) was then added and samples shaken for a further 30 secs. Samples were then centrifuged at 16,000 g for 5 min, aqueous supernatant collected and RNA precipitated with an equal volume of 4 M LiCl at 4 °C overnight. RNA was pelleted by centrifugation at 16,000 g at 4 °C for 10 mins, before resuspension in 250 µl DEPC H_2O. RNA was then precipitated by the addition of 0.1 volume of 3 M NaAcOH (pH 5.2) and 2 volumes of cold EtOH (-20 °C). Samples were vortexed and incubated at -20 °C for 1 hour before centrifugation at 16,000 g for 10 mins at 4 °C to recover the RNA pellet. RNA pellet was then washed with 500 µl of 85 % DEPC EtOH, air dried on ice and resuspended in 100 µl DEPC H_2O before storage at -80 °C.

2.2.3 cDNA synthesis

10 µg of extracted RNA (see 2.2.1) was used in a 50 µL DNase treatment reaction using a TURBO DNA-free kit (invitrogen). 1 µL treated RNA was then run on 1.5 % agarose

gel to check RNA integrity. 500 ng of treated RNA was then used in cDNA synthesis reaction using a SuperScript II reverse transcriptase kit (Invitrogen) using oligo dT18 under the following conditions: RNA and oligo dT18 denatured at 65 °C for 5 mins, chilled on ice and cDNA synthesis reagents added according to protocol. Samples then heated in a PTC225 Peltier Thermal Cycler (MJ Research) as follows: 25 °C for 10 mins, 42 °C for 60 mins and 70 °C for 10 mins.

2.2.4 qRT-PCR primer design and verification

Oligos were designed according to the MIQE guidelines (Bustin *et al.* , 2009) using the online tool Primer3 Plus, specifying a PCR amplicon between 80-130 bp, a consistent T_m for each primer pair (60-70^{o}C), primer size between 22-28 nt and a guanine-cytosine content of between 20 to 80 %. Oligos were initially tested by optimised PCR of target cDNA, and product MW analysis by agarose gel electrophoresis, staining with EtBr. Oligos were then tested for specificity by melt-curve analysis at the end of the qRT-PCR experiment (see 2.2.6), checking for a single, sharp peak representing one PCR product. The amplicon identity was then confirmed by sequencing the amplicon using LIGHTrun sequencing (GATC-Biotech). Primer efficiency was determined for each primer pair by qRT-PCR (see 2.2.6) of a standard cDNA (pooled WT cDNA) dilution series from 25 ng to $25*10^{-6}$ ng. Ct values were then plotted against Log_{10}[initial cDNA], the gradient (m) of the resulting curve calculated and efficiency (E) calculated using the formula: E = $10^{(-1/m)}$. Primer efficiency values ranging between 90-110 % were deemed acceptable for direct comparison by the $\Delta\Delta$CT method, while any primer pairs falling outside this range were corrected by the Pfaffl method (Pfaffl, 2001). All primer sequences can be found in 7.4.

2.2.5 Semi-Quantitiative PCR

25 ng cDNA was subjected to PCR analysis using gene specific primers for the gene of interest and the housekeeping gene in duplicate reactions. Each cDNA sample was additionally analysed at 3 further dilutions: 1/10, 1/20 and 1/30. Reactions were carried out in a PTC225 Peltier Thermal Cycler (MJ Research) with an initial denaturation step of 95 °C for 30 sec, followed by cycles of denaturation at 95 °C for 30 sec, annealing at a primer specific temperature for 30 sec and extension at 72 °C for 1 min per kb of product. PCR products were electrophoresed on 1xTAE agarose (Invitrogen) gels (1.5 %, w/v) and imaged on a transilluminator using ethidium bromide.

2.2.6 Quantitative Real-Time PCR

25 ng of synthesised cDNA (see 2.2.3) was analysed by qRT-PCR in triplicate 15 µL reactions containing 7.5 µL MESA blue qPCR mastermix for SYBR Green assay (Eurogentec), 0.15 µL of 20 mM forward and reverse oligos (see 2.2.4). Each sample was analysed in separate reactions using both oligos specific for the gene of interest and the ubiquitous housekeeping gene *Glyceraldehyde-3-phosphate dehydrogenase* (*Gapdh*) (gene identifier M18976). qRT-PCR reactions were carried out using a MyiQ detection system (BioRad) under the following conditions: 95 °C for 5 min followed by 45 cycles of 95 °C for 15 sec and 64 °C for 1 min. Data were analysed using iQ5 optical system software (BioRad) and Excel software (Microsoft). CT (threshold cycle) values were normalised to give ΔCT by subtracting the arithmetic mean of the reference housekeeping genes' CT values. Relative Expression (RE) values were then calculated by the formula $RE = 2^{-\Delta CT}$. Expression ratios were then calculated using the formula Expression Ratio = RE_{sample} / $RE_{control}$. Values were then Log transformed to give Log Expression Ratio and plotted using the graphing software Prism (GraphPad). All

errors associated with mean values were propagated through the data analysis to give a final propagated standard error, which was then used to generate error bars.

2.2.7 Maize Genomic DNA extraction

High throughput Leaf samples approximately 10 mm^2 were collected from each plant and stored at -80 °C in 96 well extraction blocks. The tissue was ground to a fine powder with steel ball bearings using a mixer mill. 300 µl of extraction buffer (100 mM Tris-HCl, pH 8.0; 50 mM EDTA, pH 8.0; 500 mM NaCl; 10 mM β- Mercaptoethanol) and 20 µl 20 % SDS was added, vortexed, and incubated for 20 min at 65 °C, followed by 10 min on ice. 100 µl of 5 M potassium acetate was added and samples were incubated for a further 20 min on ice. After a 13 000 rpm centrifugation, 125 µl of the supernatant was transferred to 200 µl isopropanol and incubated at -20 °C for 30 min. Following further centrifugation at 13 000 rpm, the isopropanol was removed and the pellet washed in 70 % ethanol. The pellets were then dried for 5 mins in a vacuum centrifuge at RT and then resuspended in 100 µl TE buffer (10 mM Tris-HCl; 1 mM EDTA; pH 8.0) containing 20 µg mL^{-1} RNaseA (Invitrogen). DNA was then stored at 4 °C.

Small-scale DNA extracted using essentially the same method as in 2.2.7, except tissue was ground in a cold pestle and mortar with liquid $N_2$and extraction carried out in 1.5 ml microfuge tubes.

2.2.8 Maize genomic DNA PCR reactions

PCR reactions were carried out in a volume of 50 µl, using 5 µl template DNA, 1 µl each of gene specific forward and reverse primers (20 µM stock), 0.5 µl dNTP's (10 mM stock), 1 µl DMSO, 5 µl 10x PCR buffer (inc. $MgCl_2$) (Invitrogen), 0.25 µl Taq

polymerase (Invitrogen) and 36.25 µl dH_2O. Reactions were carried out in a PTC225 Peltier Thermal Cycler (MJ Research) with an initial denaturation step of 95 °C for 30 sec, followed by cycles of denaturation at 95 °C for 30 sec, annealing at a primer specific temperature for 30 sec and extension at 72 °C for 1 min/kb of product. PCR products were electrophoresed on 1xTAE agarose (Invitrogen) gels (1.5 %, w/v) and imaged on a transilluminator using Ethidium Bromide.

2.2.9 Genotyping transgenic seeds by GFP reporter gene

In transgenic lines which contained the GFP selection marker gene, seed genotyping was carried out by examining the whole ear under UV light, selecting the fluorescent seeds as positive and marking them with a permanent marker. Seeds were then removed from the ear and the positive seeds selected.

2.2.10 Genotyping transgenic plants by molecular methods

In transgenic lines lacking the GFP selection marker gene, molecular genotyping was carried out by germinating seeds (see 2.1.1) and sampling the young leaves for DNA extraction (see 2.2.7) and PCR (see 2.2.8), using oligos specific for the marker gene of interest.

2.2.11 Generation of endosperm specific cDNA libraries

Approximately 50 endosperms per line were isolated from developing seeds at 6 DAP by micro-dissection and immersed immediately into RNAlater solution (Ambion). RNAlater was then removed and the samples frozen in liquid nitrogen before storage at -80 °C. RNA was then extracted from endosperm tissue using the hot phenol method (see 2.2.2),

DNase treated (see 2.2.3) and quantified by Nanodrop. RNA integrity was then verified by BioAnalyser and cDNA libraries generated using the TruSeq Stranded mRNA Sample Prep Kit (Illumina).

2.3 Cloning Methods

2.3.1 Synthetic gene design

***E.coli* MLP1 construct** Full length amino acid sequence for MLP1 was analysed using the web based signal peptide sequence finder SignalP 4.1 (Petersen *et al.* , 2011). The amino acid sequences were then truncated to remove the signal peptides, tagged with 6 histidine residues, back-translated and codon optimised for *E.coli* using the web based BackTranslation tool (Entellechon) (Fischer, n.d.). A start codon was also included since the native codon had been truncated. These DNA sequences were then appended to the restriction sequences for *Nde*I and *Bam*HI to allow for cloning. Construct was then synthesised by GenScript USA Inc.

***Z.mays* MLP1 construct** The full length ORF DNA sequence for *ZmMLP1* was obtained from the maize genome (GRMZM2G145466_T01), the stop codon removed, and used as the base for the construct. To allow down-stream removal of the His tag, the PreScission protease recognition sequence ProGly/GlnPheLeuValGluLeu (Cordingley *et al.* , 1990) was back translated, maize optimised (Fischer, n.d.) and added to the 3' end of the ORF. The maize optimised (Fischer, n.d.) DNA sequence for the hexahistidine tag was then added to the 3' end of the PreScission sequence, followed by a stop codon. To facilitate use of the Gateway (Invitrogen) cloning system, the attb1 and reverse-complimented attb2 (using SeqBuilder, DNASTAR) were appended to the 5' and 3' ends of the construct. Construct was then synthesised by GenScript USA Inc.

2.3.2 Cloning synthetic genes into pET29a expression vector

Synthetic genes in pUC57 vectors were transformed into competent *E.coli* using a heat-shock protocol (see 2.3.3). Positive transformants were picked and cultured O/N at 37 °C in 5 ml LB + Kanamycin. Cells were then pelleted and the plasmid DNA extracted using the QIAprep MiniPrep kit (QIAGEN). 4 µg plasmid was then digested by both *Bam*H1 and *Nde*1 restriction enzymes at 37 °C for 1 h. Digest products were then run on a 1.8 % agarose gel and the correct size fragments containing the synthetic gene were excised from the gel under UV light. DNA was then purified from the gel fragment using the QIAquick gel extraction kit (QIAGEN) and a sample run on a 1.8% agarose gel to check purity. DNA fragments were then ligated into the pET29a plasmid at 4 °C using the T4 DNA ligase enzyme.

2.3.3 Transformation of DH5α competent *E.coli* cells

40 µl of thawed DH5α competent *E.*coli *cells* were mixed with 4 µl of purified plasmid DNA and cooled on ice for 30 mins. Cells were then heat-shocked at 42 °C for 1 min then returned to ice for 3 mins. 500 µl SOB media was added to cells followed by incubation at 37 °C for 1 h. Cells were then grown on LB agar media + vector antibiotic overnight at 37 °C.

2.3.4 Identification of positive transformants by colony PCR

Colonies were picked and transferred to 5 µl dH_2O and a replica plate was made. A standard PCR mix was made up using T7e (fwd) and pET32 (rev) primers to a total volume of 30 µl. A positive control was included containing pET29a plasmid DNA. Samples were reacted in a PTC225 Peltier Thermal Cycler (MJ Research) in the following cycle: 95 °C 2 mins, 95 °C 30 sec, 56 °C 30 sec, 72 °C 20 sec for 25 cycles,

ending in 72 °C for 10 mins then 10 °C ∞. PCR products then run on a 2 % agarose gel.

2.3.5 Sequencing of positive transformants

Positive colonies from duplicate plate were picked and cultured in LB Kanamycin. Cells were then pelleted and the plasmid DNA extracted using the QIAprep MiniPrep kit (QIAGEN). DNA was then sequenced using the Sanger method with in-house sequencing.

2.3.6 Transformation into Rosetta 2 DE3 cells

25 µl of thawed Rosetta 2 DE3 pLysS *E.coli* cells (Invitrogen) were mixed with 25 ng of plasmid DNA and transformation carried out as in 2.3.3. Cells were grown on LB agar Kanamycin (0.2 mg mL^{-1}) Chloramphenicol (0.017 mg mL^{-1}) at 37 °C O/N. 3 colonies of each transformed cell line were picked, cultured in LB Kanamycin (0.2 mg mL^{-1}) Chloramphenicol (0.017 mg mL^{-1}) at 37 °C until OD_{600}reached 0.6. Glycerol stocks were then made using 180 µl 80 % sterile glycerol and 820 µl liquid cell culture, mixed, frozen using liquid $N_2$and stored at -80 °C.

2.3.7 Gateway Cloning

Specific forward and reverse oligonucleotides were designed for the gene of interest, fusing the attB1/attB2 sequences (Invitrogen) with the first 21 5' nucleotides of the gene of interest. The gene of interest was then amplified by PCR using the high fidelity AmpliTaq Gold® 360 Master Mix protocol (Applied Biosystems) under the following conditions: 95 °C 10 mins, 95 °C 15 sec, 45 °C 15 sec, 72 °C 2 mins for 5 cycles, followed by: 95 °C 15 sec, 55 °C 15 sec, 72 °C 2 mins for 35 cycles, ending in 72 °C for 5 mins then

10 °C ∞. PCR products were then run on a 1.5 % agarose gel to confirm the amplicon size followed by PCR clean-up using the QIAquick PCR purification kit (Qiagen). For synthetic genes, the construct was designed to include the attB1 and attB2 sequences flanking the coding sequence, therefore no PCR amplification was required. attB flanked DNA (PCR product or plasmid) was then cloned into a DONR vector plasmid by BP recombination reaction according to the BP clonase kit protocol (Invitrogen). BP reaction product was then transformed into DH5α by heat-shock (see section 2.3.3). Colonies were picked and grown in liquid LB media + vector antibiotic O/N at 37 ºC before plasmid extraction by QIAprep MiniPrep protocol (QIAGEN). Plasmid DNA was then quantified by Nanodrop and 150 ng cloned into a Gateway compatible binary vector for plant expression, according to the LR clonase kit protocol (Invitrogen). LR recombination product was then transformed into DH5α by heat-shock (see section 2.3.3). Colonies were picked and grown in liquid LB media + vector antibiotic O/N at 37 ºC before plasmid extraction by QIAprep MiniPrep protocol (QIAGEN). Plasmid DNA was then quantified by Nanodrop.

2.3.8 *At*ESF1.3/*Zm*MLP1 domain substitution constructs

Constructs were designed by substituting domains of interest from the *Zm*MLP1 amino acid sequence in place of the corresponding domain in the full length *At*ESF3 amino acid sequence (see 2.6.2) while avoiding secondary structure, conserved aromatics and cysteine residues. These sequences were then tagged between the signal peptide sequence (predicted by Signal P 4.0 server (Petersen *et al.* , 2011)) at the N-terminus with a FLAG tag sequence (DYKDDDDK). The protein sequences were then back translated and *Arabidopsis thaliana* optimised (Fischer, n.d.). Sequences were then flanked with Gateway attB1/2 sequences (Invitrogen) and synthesised in pUC57 plasmids (GenScript USA Inc). Constructs were then cloned by Gateway (see 2.3.7) into

the plant over-expression vector pMDC ON2.1 (Costa *et al.* , 2014) which was generated by digesting the pOpON2.1 vector (Samalova *et al.* , 2005) with the restriction enzyme NotI and ligating the fragment into pMDC123 (Curtis & Grossniklaus, 2003).

2.4 Biochemical Methods

2.4.1 Protein expression in Rosetta *E.coli*

Small scale expression Colonies were grown on LB agar Kanamycin (0.2 mg mL^{-1}) Chloramphenicol (0.017 mg mL^{-1}), picked and cultured in 5 ml LB Kanamycin (0.2 mg mL^{-1}) Chloramphenicol (0.017 mg mL^{-1}) at 28 °C O/N. This starter culture was then used to inoculate 5 ml LB media without antibiotics which was grown at 37 °C until OD_{600}reached 0.6. Expression was then induced using IPTG (concentration optimised) and the culture grown for 4 h at 37 °C (optimised). 1 ml of cell culture was then spun at 16 000 g for 1 min and the supernatant discarded.

Up-scaled expression Starter culture generated as in 2.4.1 and used to inoculate 50 ml LB media which was grown at 37 °C until OD_{600}reached 0.6. Expression was then induced with IPTG (concentration optimised) and culture allowed to grow for 4 h at 37 °C. 1 ml culture then sampled, spun at 16 000 g and supernatant discarded. Pellet boiled in 1x SDS Laemmli buffer and checked for expression by SDS-PAGE (see 2.4.6). Remaining culture pelleted at 16 000 g, supernatant discarded and stored at -20 °C.

2.4.2 Protein extraction from maize seeds

Maize kernels were harvested and bisected with a razor blade, separating the antipodal (top) and micropylar (bottom) halves of the seed. Pooled halves were then ground in liquid $N_2$and the proteins extracted by TCA NaDOC precipitation (see2.4.3).

2.4.3 Protein extraction by TCA NaDOC precipitation

A crude protein extraction was prepared by tissue disruption in four volumes (w/v) of cold PBS with protease inhibitors and spun at 16000 g for 10 mins to remove cellular debris. The supernatant was then precipitated on ice by addition of 0.01 volume of 2 % NaDOC (sodium deoxycholate) followed by 0.1 volume of 100 % (w/v) TCA (trichloroacetic acid). Proteins were then pelleted by centrifugation at 16,000 g for 10 mins at 4 ^{o}C, washed twice in 500 µL ice-cold acetone and dissolved in 100 µL 1x SDS Laemmli buffer at 95 ^{o}C for 10 mins.

2.4.4 Protein extraction by acetone precipitation

A crude protein extraction was prepared by tissue disruption in four volumes (w/v) of cold PBS with protease inhibitors and spun at 16000 g for 10 mins to remove cellular debris. Crude protein extract was then added to four times the sample volume of cold acetone at -20 °C and incubated at -20 °C for at least 1 h. Sample was then centrifuged at 16 000 g for 10 mins and supernatant discarded. The protein pellet was then washed 3 times in cold 80 % acetone, and allowed to air dry for 30 mins on ice before resuspension in an appropriate buffer.

2.4.5 Protein quantification

Protein samples were quantified by one of the following methods: The protein sequence was used to calculate the extinction coefficient (ε) with the web-based tool ProtParam (Elisabeth Gasteiger, 2005). The UV absorbance of the protein at 280 nm was measured after blanking with buffer, and using the Beer-Lambert law, the $A_{280}$and ε were used to calculate the protein concentration. If however, the sequence lacks tryptophan residues, ε is unreliable. Therefore, the protein was quantified using the Bradford

assay (Biorad). A standard curve was generated using BSA and the sample protein concentration estimated.

2.4.6 SDS-PAGE of proteins

Unless otherwise stated, protein samples were mixed with SDS Laemmli buffer (Laemmli, 1970), boiled for 10 mins and loaded onto a 5 % stacking, 15 % resolving tris-glycine SDS polyaccrylamide gel, using the mini-Protean system (BioRad). Spectra Multicolour Low Range Protein Ladder (Fermentas) was also loaded and proteins were then electrophoresed at 200 V until dye front reached gel bottom. Proteins were then visualised using Coomassie R250 dye or silver nitrate.

2.4.7 Tricine-SDS PAGE of proteins

Proteins requiring high resolution at low MW were separated by Tricine-SDS-PAGE essentially using the method described in (Schägger, 2006). Proteins were then visualised using Coomassie R250 dye or silver nitrate.

2.4.8 Protein purification by metal affinity

Cell pellet from protein expression (see 2.4.1) was lysed and purified using the QIAexpressionist mini-prep protocols under both native and denaturing conditions with nickel agarose NTA beads (QIAGEN). Eluted protein was then electrophoresed (see 2.4.6) to check purity and sequenced by mass spectrometry to confirm identity (see 2.4.10).

2.4.9 Protein purification by Immobilised Metal Affinity Chromatography (IMAC)

Cell pellets from large scale expression (see 2.4.1) were lysed in buffer A (50 mM $NaPO_4$, 300 mM NaCl, pH 8.0) using lysozyme at 1 mg mL^{-1} for 30 mins at RT followed by sonication for 1 min. Lysate cleared by centrifugation at 16 000 g for 30 mins at 4 °C. Supernatant was then purified by IMAC on an Akta purifier (GE Healthcare) as follows. A 5 ml nickel-sepharose column (GE Healthcare) was stripped and regenerated with nickel chloride. Column was then equilibrated with buffer A before clarified lysate was injected via a super-loop. Column was washed with buffer A until OD_{280}flatlined. Protein was then eluted into a fraction collector by increasing concentration of buffer B (50 mM $NaPO_4$, 300 mM NaCl, 500 mM imidazole, pH 8.0) stepwise. Fractions were analysed by SDS-PAGE (see 2.4.6), before pooling pure fractions and buffer exchange into 20 mM PO_4pH 7.3 using a Vivaspin 20 PES 3000 MWCO column (Sartorius Stedim).

2.4.10 Protein identification by Mass Spectrometry

The Coomassie stained gel pieces were processed and tryptically digested using the manufacturer's recommended protocol on the MassPrep robotic protein handling system. The extracted peptides from each sample were analysed by means of nanoLC-ESI-MS/MS using the NanoAcquity/Synapt HDMS instrumentation (Waters) using a 45 minute LC gradient. All MS and MS/MS data were corrected for mass drift using reference data collected from the [Glu1]-Fibrinopeptide B (human - F3261 Sigma) sampled each minute of data collection. The data were used to interrogate the E. coli MC4100 database (http://www.ebi.ac.uk/integr8) appended with the sequence supplied using ProteinLynx Global Server v2.4.

2.4.11 Antibody Production

800 µg of purified MLP1 was sent to Eurogentec (Belgium) and used to immunise 2 rabbits, using 4 injections of 100 mg per rabbit across a 3-week period. Pre-immune, medium and final serum bleeds were taken before, 3 and 4 weeks after initial immunisation.

2.4.12 Enumeration of disulphide bonds in protein by Ellman's reaction

Protein sample was quantified (see 2.4.5) and two aliquots (one to be reduced, one oxidised) made up to 20 µM in 20 mM PO_4, 10 M urea, 1 mM DTT (Reduced sample only), pH 8.5 and incubated for 30 mins at RT. Samples then buffer exchanged into 10 M urea, 100 mM Tris pH 8.5 on a pre-equilibrated PD-10 desalting column. Samples then concentrated using a MWCO column and quantified (see 2.4.5). 250 µl of reduced and oxidised protein then made up to 1 ml in 10 M Urea, 1mM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB), 100 mM Tris pH 8.5 along with a range of standard concentrations of reduced glutathione (GSH) (0 - 100 µM). Standards and samples incubated at RT for 20 mins and absorbance measured at 412 nm. Standard curve generated and used to estimate the concentration of free thiols in both the reduced and oxidised samples. Number (N) of reduced cysteine residues then calculated for reduced and oxidised protein by N = $[Cys_{red}]/[protein]$.

2.4.13 MLP1 secondary structure assessment by circular dichroism

Purified protein (see 2.4.9) was made up to an optimised concentration to keep High Tension (HT) signal below 500 V and the CD signal above -10 md. Sample and buffer analysed by circular dichroism using 1 mm quartz cuvette, at a range of 260 - 180 nm, at 0.5 nm steps. Sample spectra buffer-subtracted and submitted to DichroWeb decon-

volution server using the CDSSTR method (Sreerama & Woody, 2000) and reference database SP175 (Lees *et al.* , 2006).

2.4.14 pro*Mlp1*-GUS expression and detection

Transgenic maize seed containing a pro*Mlp1*:GUS fusion gene (see 2.1.7) were grown to maturity under glasshouse conditions (see 2.1.1), sib-crossed (see 2.1.5) and the progeny sampled at both pre-fertilisation and 10 DAP. Sampled seeds were pre-cut, washed in 100 mM $NaPO_4$, pH 7.0 buffer and incubated in GUS stain solution (50 mM $NaPO_4$, pH 7.0; 2 mM $K_3[Fe(CN)_6]$; 2 mM $K_4[Fe(CN)_6]$; 0.1 % Triton X-100; 2 mM X-GlucA). 10 DAP samples were vacuum infiltrated and stained at 37 ^{o}C overnight, while pre-fertilisation samples were stained at 28 ^{o}C overnight. Samples were then fixed in formaldehyde (see 2.5.1), wax embedded and sectioned (see 2.5.2). Whole samples (pre-embedding) were photographed using a Zeiss Discovery V12 stereo microscope with an AxioCam digital camera. Sectioned samples were photographed using a Zeiss AxioSkop microscope equipped with a Nikon DXM 1200F digital camera.

2.4.15 Immunodetection of proteins

Protein samples were separated by SDS-PAGE (see 2.4.6) and transferred to Immobilon-P^{SQ} PVDF membrane (Millipore) by Western blotting. Blot carried out using semi-dry blotting module (BioRad) at 12 V, 80 mA for 30 mins. Membrane was then blocked using 5 % skimmed milk in PBS for 30 mins at RT. The blot was then incubated in antigen-specific primary antibody in 3 % BSA, PBS at an optimised concentration O/N at 4 °C. Blot then washed 3 times in PBS 0.1 % Tween 20 for 15 mins followed by incubation with a secondary antibody raised against the primary antibody host animal for 1 h at RT at an optimised concentration. Blot was then washed 4 x 15 mins in

PBS 0.1 % Tween 20 followed by 10 mins in PBS. In the case of HRP conjugated secondary antibodies, detection of secondary antibody then carried out by enhanced chemiluminescence (ECL) using SuperSignal West Pico Chemiluminescent Substrate kit (Thermo) and imaged using the ImageQuant LAS4000 (GE Healthcare). When alkaline phosphatase (AP) conjugated secondary antibodies were used, detection was carried out using SigmaFAST NBT/BCIP synthetic AP substrate (Sigma) according to manufacturers protocol.

2.4.16 Extraction of apoplastic and total protein from *N.benthamiana* leaves

Apoplastic proteins were extracted according to the method found in (Lohaus *et al.* , 2001) with the following alterations: PBS was used as the wash buffer and infiltration of leaves with wash buffer was carried out in open beakers subjected to high vacuum in a desiccator. Total proteins were extracted from the apoplast depleted leaves by grinding the leaf tissue in liquid N_2followed by 1 g of leaf powder in 2 mL QB buffer (100 mM KPO_4 pH 7.8, 1 mM EDTA pH 8.0, 1 % Triton X-100, 10 % glycerol, 1 mM DTT). 1 mL of suspension was then centrifuged at 16,000 g for 10 mins at 4 ^{o}C and the supernatant collected. Proteins from both fractions were then extracted on ice by TCA NaDOC precipitation (see 2.4.3).

2.4.17 Immunoprecipitation of GFP fusion proteins

Isolated maize endosperms from transgenic maize lines were homogenised on ice using a tissue lyser in 4 volumes (w/v) of IP lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA pH 8.0, Roche cocktail inhibitor). Samples centrifuged at 16,000 g for 10 mins at 4 ^{o}C and the supernatant collected. Protein concentration was determined by Bradford assay (see 2.4.5) and adjusted to 2 mg mL^{-1}. GFP proteins

then immunoprecipitated using washed GFP-trap agarose beads (Chromotek) for 1.5 hours at 4 ^{o}C under constant mixing. GFP-trap beads were then collected by pulse centrifugation and washed twice with IP lysis buffer before elution into 50 µL of 2x SDS Laemmli buffer at 95 ^{o}C for 10 mins.

2.4.18 Protein expression in BY2 cell suspension culture

Protein expression was carried out by our collaborators Ohki *et al* and Mori *et al* in BY2 cell suspension culture essentially as described in (Ohki *et al.* , 2008).

2.4.19 Protein purification and structural NMR

Expressed *Zm*MLP1 was purified by our collaborators Ohki *et al* and Mori *et al* by affinity chromatography and HPLC before analysis by solution state NMR, essentially as described in (Ohki *et al.* , 2011; Costa *et al.* , 2014).

2.5 Histological and Microscopic Methods

2.5.1 Maize kernel fixation

Fresh formalin-acetic-acid-alcohol (FAA) fixative was prepared as follows: 50 % Ethanol, 35 % dH_2O, 10 % Formaldehyde (Histological grade, 37 wt. % in dH_2O (Sigma)), 5 % Glacial acetic acid. Kernels were pre-cut to preserve tissue of interest, placed in sample cassettes and immersed in cold FAA. Fixed material was then exposed to a moderate vacuum for 10 mins, slowly re-pressurised, and FAA changed. Tissue then fixed O/N at 4 ^{o}C before dehydrating in an ethanol series (50 % - 70 %) and storage at 4 ^{o}C.

2.5.2 Tissue sectioning

Plant tissue was fixed as in 2.5.1 followed by further dehydration, tissue clearing by Histoclear (Fisher) and embedding in paraffin wax. Embedded tissue was then sectioned at 12 μm increments using a microtome. Sections of interest were then floated on dH_2O on Superfrost Plus (Fisher) microscope slides at 37 °C. Sections were then mounted by removal of the dH_2O and incubation at 37 °C for 24 h.

2.5.3 Immunohistochemistry

Mounted formalin fixed, paraffin embedded (FFPE) tissue sections (see 2.5.2) were dewaxed twice with Histoclear tissue clearing agent (Fisher) for 10 mins. Tissue was then rehydrated in a 100-30 % EtOH series followed by a dH_2O rinse and equilibration in PBS for 20 mins. Sections were then digested by proteinase K (Sigma) for 10 mins at 37 °C, washed in PBS three times for 5 mins and incubated in PBS 0.1 % BSA for 30 mins at RT. Sections then blocked using 2.5 % goat serum in PBS/BSA for 2 hours at RT and washed for 15 mins in PBS/BSA. Sections then incubated with the primary antibody overnight at 4 °C at an optimised dilution in PBS/BSA. For alkaline phosphatase (AP) conjugated secondary antibodies, slides were washed in TBS 0.1 % BSA twice for 15 mins followed by incubation in AP conjugated secondary antibody raised against the primary antibody host organism for 2 hours at RT at an optimised dilution in TBS/BSA. Sections then washed twice in TBS/BSA for 15 mins, in TBS for 5 mins and then substrate buffer (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$) for 5 mins. Detection then carried out using NBT/BCIP AP substrate (Sigma) for 25 mins in the dark before washing in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0), dehydration through EtOH series (30-100 %) and mounting under a cover slip using Entellan mountant. For fluorescein isothiocyanate (FITC) and tetramethyl

rhodamine iso-thiocyanate (TRITC) conjugated secondary antibodies, 3 washes carried out after primary antibody incubation were performed in PBS 0.1 % Tween followed by a rinse in PBS. Sections were then incubated in secondary antibody for 2 h at RT in the dark followed by a further 3 washes in PBS 0.1 % Tween and a rinse in PBS all carried out in the dark. Sections were then mounted using VECTASHIELD mounting media (Vector), coverslipped and stored at 4 °C in the dark.

2.5.4 *In situ* hybridisation

Generation of DIG-labelled RNA probes

All solutions used during RNA probe labelling were RNase free. RNA probes were generated by amplification of cDNA (see 2.2.3) using oligos specific for the gene of interest, and cloning into a T7 plasmid. The plasmid was then transformed into *E.coli* (see 2.3.3), propagated in culture and extracted by Miniprep (Qiagen) before liniarisation using restriction enzymes. RNA probe was then generated and labelled by RNA polymerase reaction in a total volume of 20 µl, using ~1 µg DNA, 1 mM ATP, CTP and GTP, 0.7 mM UTP and 0.35 mM UTP-DIG, 1 x transcription buffer (Promega), 20 mM DTT, 40 units RNase OUT (Invitrogen) and 20 units of T7 polymerase at 37 °C for 1 h. RNA was then Turbo-DNase treated (Ambion) and RNA precipitated using 0.5 vols 7.5 M ammonium acetate and 4 vols ETOH at -20 °C O/N. Precipitated RNA was then centrifuged at 16 000 g for 20 mins at 4 °C. The pellet was then washed with 85 % ETOH and centrifuged as before. The supernatant was discarded and the pellet air-dried, re-suspended in 50 µl dH_2O and hydrolysed at 60 °C with 100 mM Na_2CO_3 pH 10.2 (probe-length optimised duration). RNA was precipitated with 0.2 % HAc, 83 mM NaAc and 72 % ETOH at -80 °C for 1 h followed by centrifugation at 16 000 g at 4 °C, washed with 85 % ETOH and air-dried. The pellet was re-suspended in 50 µl

dH_2O and efficiency of the labeling was checked using a dot blot with the remainder being stored in 50 % deionised formamide (Sigma).

RNA *in situ* Hybridisation.

All solutions used during tissue preparation and hybridisation steps were RNase free. Mounted formalin fixed, paraffin embedded (FFPE) tissue sections were de-waxed twice in Histochoice (AMRESCO) for 10 mins. Tissue was then rehydrated through a 100-30 % ETOH series followed by equilibration in dH_2O and PBS for 2 mins each. Sections were then incubated in 0.125 mg ml^{-1} Pronase (Sigma) in PBS for 10 mins at 37 °C, followed by 2 mins in 0.2 % (v/v) glycine in PBS washed in PBS. Sections were then immersed in 4 % (w/v) formaldehyde in PBS for 10 mins, followed 2 x 2 mins washes in PBS. Finally, the sections were incubated in 0.5 % (v/v) acetic anhydride in 0.1 M triethanolamine pH 8 for 10 mins, washed in PBS for 2 mins, dehydrated back through ETOH series and air-dried. RNA probe was denatured at 80 °C for 2 mins before being added to hybridisation buffer (0.3 M NaCl, 10 mM Tris-HCl pH 6.8, 10 mM $NaPO_4$, 5 mM EDTA, 2 % (v/v) deionised formamide, 1.25 mg ml^{-1} tRNA, 2.5 x Denhardt solution (Sigma) 12.5 % (v/v) dextran sulfate (Sigma)) in a 1:4 ratio. 50 µl of probe-buffer solution was added to each slide and covered with a Sigmacote (Sigma) treated coverslip and hybridised O/N at an optimised temperature. The sections were then washed in 0.2 % (v/v) saline-sodium citrate buffer (SSC) to remove coverslip and washed twice in fresh 0.2 % (v/v) SSC at 55 °C for 30 and 90 mins respectively. Sections were then washed in NaCl-Tris-EDTA buffer (NTE) twice for 5 mins at 37 °C and then incubated in 20 µg ml^{-1} RNaseA (Sigma) for 30 mins at 37 °C. The sections were then washed 3 x 5 mins in NTE, washed in 0.2 % (v/v) SSC for 1 h and finally washed in PBS for 5 mins. Slides were incubated in DIG-buffer 1 (1 M Tris-HCl, 1.5 M NaCl pH 7.5) for 5 mins before incubation in DIG-buffer 2 (0.5 % (w/v) blocking reagent

(Roche) in DIG-buffer 1) for 30 mins under gentle agitation. 100 µl of DIG-buffer 3 (1 % (w/v) BSA, 0.3 % (v/v) triton-x-100 in DIG-buffer 1) was then added to each slide and covered with a plastic coverslip and incubated for 30 mins. Coverslips were then removed, slides drained and 100 µl of 1:3000 dilution of Anti-digoxigenin-AP antibody (Roche) in DIG-buffer 3 was added, incubating for 1.5 h. Slides were then washed 4 x 20 mins in 0.3 % (v/v) triton-x-100 in DIG-buffer 1, followed by DIG-buffer 1 for 10 mins and finally DIG-buffer 5 (100 mM Tris (pH 9.5), 100 mM NaCl and 50 mM $MgCl_2$) for 5 mins. Slides were then incubated in SigmaFast BCIP/NBT (Sigma) overnight in the dark. Reaction was terminated in Tris-EDTA buffer for 5 mins, slides washed in dH_2O for 2 mins and dehydrated through 30-100 % ETOH series before rehydration back through EOTH series and finally into dH_2O. Slides were then air-dried and mounted in Entellan (MERCK).

2.5.5 Laser scanning confocal microscopy of GFP fusion proteins

Tobacco leaves Infiltrated tobacco leaf tissue (see 2.1.11) was cut into small squares and mounted on a microscope slide using water, with the lower epidermis facing up. For plasmolysis experiments, leaves were incubated in plasmolysing solution (1 M mannitol, 0.01 % Tween 20) for several hours, until plasmolysis observed. GFP expression was then analysed using a Carl Zeiss LSM 710 inverted confocal laser scanning microscope. GFP was excited at 488 nm and the fluorescence emission was detected at 505-530 nm. Image analysis was carried out using Zeiss Efficient Navigation (ZEN) software.

Maize endosperm Intact maize endosperms from developing maize seeds were micro-dissected out and mounted ESR side up in water on a microscope slide and cover slipped. GFP expression was then analysed using a Carl Zeiss LSM 710 inverted confocal laser scanning microscope. GFP was excited at 488 nm and the fluorescence emission was

detected at 505-530 nm. Image analysis was carried out using Zeiss Efficient Navigation (ZEN) software.

Arabidopsis seeds

Developing seeds were removed from the plant and mounted on microscope slides in water. GFP expression in endosperms was analysed using a Carl Zeiss LSM 710 inverted confocal laser scanning microscope equipped with a UV lamp and digital camera. Image analysis was carried out using Zeiss Efficient Navigation (ZEN) software.

2.5.6 Epifluorescence microscopy of GFP fusion proteins

GFP expression in developing maize endosperms and ovules was analysed using a Carl Zeiss LSM 710 inverted confocal laser scanning microscope equipped with a UV lamp and digital camera. Image analysis was carried out using Zeiss Efficient Navigation (ZEN) software.

2.5.7 Internal seed morphology analysis

MLP1 RNAi seeds were qualitatively selected according to exhibition of an external phenotype. These seeds, together with WT controls (RNAi -ve) were removed from the ear and imbibed in water for 12 h before bisection along the central seed axis with a razor blade. Seeds were then either analysed directly or immersed in iodine solution (Gram iodine) under gentle vacuum for 10 mins to stain starch. Developing seeds were processed in the same way. Seeds were then oriented with the internal side facing down on a petri dish and imaged by scanning on a document scanner.

2.5.8 Histological staining

Toluidine Blue O staining

FFPE tissue sections were de-waxed twice with Histochoice (AMRESCO) for 10 mins and rehydrated in a 100-30 % ETOH series and subsequently into dH_2O. Sections were stained in 0.05 % (w/v) Toluidine Blue O (SIGMA) solution for 1.5 mins before being washed in dH_2O for 30 secs. Samples were washed 2 x 5 mins in dH_2O. Sections were then dehydrated back through above ETOH series and air-dried. Slides were then mounted in Entellan (MERCK), cover slipped and stored at RT.

2.6 *In silico* Methods

2.6.1 Amino acid sequence alignments

Amino acid sequences were aligned using Jalview(Waterhouse & Barton, 2009) using the ClustalW algorithm(Thompson *et al.* , 1994). Aligned residues were then colour coded by residue with the Taylor colour scheme (Taylor, 1997) or by physio-chemical properties using the Zappo scheme (Waterhouse & Barton, 2009). In the case of cysteine anchored alignments, the residue coordinates for the anchor points were submitted along with the amino acid sequences to the DIALIGN server (Morgenstern *et al.* , 2006) and the resulting alignment file imported into Jalview. Sequence similarity was assessed by submission of target sequences to the SIM online tool (ExPASy) and analysis using the BLOSUM62 algorithm (Huang & Miller, 1991).

2.6.2 Identification of *At*ESF3/*Zm*MLP1 structural loops

Preliminary structural investigation was carried out by means of *In silico* secondary structure prediction using the Jpred server (Cole *et al.* , 2008) and 3D structure ho-

mology modelling. Upon obtaining the NMR structure of *Zm*MLP1, the secondary structure was aligned with the NMR data for *At*ESF3 (see 2.6.1) together with the disulphide bond pairings. Both conventional Clustal and anchored alignments were performed and annotated for clarity before analysis together with the 3D NMR model to identify inter-disulphide loop regions.

2.6.3 Promoter Region Sequence Analysis

The promoter regions of genes of interest were obtained from online databases (NCBI, Maize GDB) using the gene accession number. 520 bp of the 5' flanking sequence starting from the ATG start codon of ESR specific genes were aligned using the ClustalWS algorithm in Jalview and regions of homology located. Potential TATA box elements were identified by sequence search.

2.6.4 Analysis of protein biophysical properties

In order to obtain biophysical data, protein sequences of interest were submitted to the ProtParam online tool (ExPASy) for analysis (Walker, 2005).

2.7 Phenotype analysis methods

2.7.1 Measurement of dry seed area

Dry seeds were extracted from ears and photographed next to a scale bar at a constant magnification using a digital SLR camera. Images were then converted to 8 bit and digitally thresholded using Image J to remove all background. Remaining binary images were then analysed using the particle analysis tool on Image J, calibrated to the photographed scale bar, yielding area measurements for each individual seed.

2.7.2 Dry seed weight measurements

Seeds from lines of interest were removed from the central portion of the ear and weighed to an accuracy of 0.001 g on a chemical balance. Individual seed weights were recorded and data analysis carried out using statistical methods.

2.7.3 Measurement of starch proportional area

Developing seeds at 15 DAP were bisected across the seed axis and immersed in Gram iodine solution for 10 mins under gentle vacuum. Dry seeds were imbibed in water for 12 h before bisection and staining. Areas of the seed containing a high concentration of starch were stained black. Seeds were then placed stained-side down into a large petri dish with a scale bar and scanned using a document scanner. Images were then converted to 8 bit, duplicated and one image digitally thresholded using Image J to remove all unstained areas. The other image was thresholded to remove only the background, leaving the total seed cross-sectional area. The binary images of the stained area were then analysed using the particle analysis tool on Image J, calibrated to the photographed scale bar, and mapping back onto the binary image of the total area, yielding % area of seed stained measurements for each individual seed.

2.7.4 Dimensional analysis of developing maize embryos

Seed sections subjected to *in situ* hybridisation or histological staining were photographed using a microscope and digital camera, calibrating the image with a graticule. Images were then analysed using ImageJ to quantify proembryo length, suspensor length, embryo proper area and suspensor area.

2.7.5 Measurement of *Arabidopsis* suspensor length

Siliques were harvested from *Arabidopsis* plants at 3 DAP, the developing seeds manually dissected out and immersed in water. Confocal microscopy was then used to image the embryo suspensors (see 2.5.5), exciting at the GFP wavelength. Images were then analysed with ImageJ to measure the length of the suspensors across multiple transgenic lines and wild type controls.

2.8 Plant tissue culture

2.8.1 Calli induction and maintenance

Maize seeds were sterilised by washing in 80 % EtOH for 3 mins, wash 50 % bleach for 15 mins, EtOH and bleach washed repeated, wash in sterile dH_2O five times. Seeds were then left in stirring sterile dH_2O for 24 h to imbibe. Seeds then transferred to solid media plates (4.4 g L^{-1} Murashige and Skoog (MS), 30 g L^{-1} sucrose, 9 g L^{-1} agar, pH 5.8) and allowed to germinate at 25 °C in the dark. Root explants were then transferred to MS, sucrose, agar media containing the synthetic growth regulator 2,4-Dichlorophenoxyacetic acid (2,4-D) at 2 mg L^{-1} under sterile conditions. Plates then incubated at 25 °C for 2 weeks in the dark. After formation, the calli were transferred to fresh MS, sucrose, 2,4-D, agar media every 2 weeks under sterile conditions.

2.8.2 Liquid plant cell suspension

A small amount of calli material (see 2.8.1) was cut into small pieces and transferred to 50 mL liquid MS, sucrose, 2,4-D media under sterile conditions and incubated with agitation at 25 °C in the dark. Sub-culturing was carried out every week by inoculation of fresh media with 10 ml cell suspension.

2.8.3 Expression induction with dexamethasone (DEX)

Solid calli tissue or liquid cell suspension was incubated in media including 50 µM DEX under optimised conditions.

2.9 Statistical Methods

2.9.1 Welches Students t-test

Statistical interrogation by Students t-test was carried out in Prism (GraphPad) software, using the Welches correction where appropriate using the following equation:

$$t = \frac{\bar{\chi}_1 - \bar{\chi}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Where $\bar{\chi}_i$ = sample mean, s_i^2 = sample variance and N_1 = sample size.

2.9.2 ANOVA multiple comparisons test

One-way analysis of variance (ANOVA) with multiple comparisons (Sidak) was carried out in Prism (GraphPad). The Sidak equation was used to correct p values as follows:

$$Sp = 1 - (1 - p)k$$

Where k = number of comparisons, Sp = Sidak p-value, p = unadjusted p-value

2.9.3 Chi-squared test

Chi-squared test was used to detect statistically significant segregation amongst natural variation in Microsoft Excel using the following formula:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Where O = Observed frequency, E = Expected frequency

3 Discovery and Molecular Characterisation of the MEG-Like Protein 1 Gene

3.1 Introduction

Previous published work has shown that in maize, the imprinted gene *Maternally Expressed Gene1* (*ZmMeg1*) (GRMZM2G354335) is expressed exclusively in the Basal Endosperm Transfer Layer (BETL) of the developing kernel (Gutierrez-Marcos *et al.* , 2004). The gene product, MEG1 is a small CRP with a pivotal role in sugar partitioning within the seed through regulation of BETL transfer cell development (Costa *et al.* , 2012).

ZmMeg1 is a member of the *Meg1* family that forms a gene cluster on maize chromosome 7, identified by screening an endosperm cDNA library constructed from maize endosperm tissue at 7 DAP (Gutierrez-Marcos *et al.* , 2004; Hueros *et al.* , 1999a). Further analysis by screening subtracted cDNA libraries constructed from specific cell types, identified another gene on the periphery of the *Meg1* cluster with homology at the protein level, but differences in the promoter region (Le *et al.* , 2005). From identified novel transcripts, homology searches highlighted an expressed sequence tag (EST) in the central cell, which has been termed *Meg like protein1* (*ZmMlp1*) (GRMZM2G145466_T01).

Previous chromosomal alignment work analysing maize chromosome 7 and 2, with sorghum chromosome 2, shows that while there have been many chromosomal rearrangements, there is a high level of cluster conservation (Figure 3.1). The *ZmMeg1* family can be found on maize chromosome 7, but has been lost from maize chromosome 2. *ZmMlp1* has been mapped to the end of the *ZmMeg1* cluster and interestingly, is the only gene conserved in sorghum 2 (Figure 3.1). Further homology searches in the genomes of barley, wheat and rice found homologous genes for *ZmMlp1*, but not for *ZmMeg1*. This conservation pattern could be indicative of an ancestral origin for *ZmMlp1*, giving rise to the *ZmMeg1* family through multiple gene duplication events in maize. Though *ZmMeg1* has been extensively characterised, nothing is known about the expression

pattern, gene product or function of *ZmMlp1*. As a first step towards identifying the function of *ZmMlp1*, the aim of this chapter is to characterise its expression in *Zea mays*.

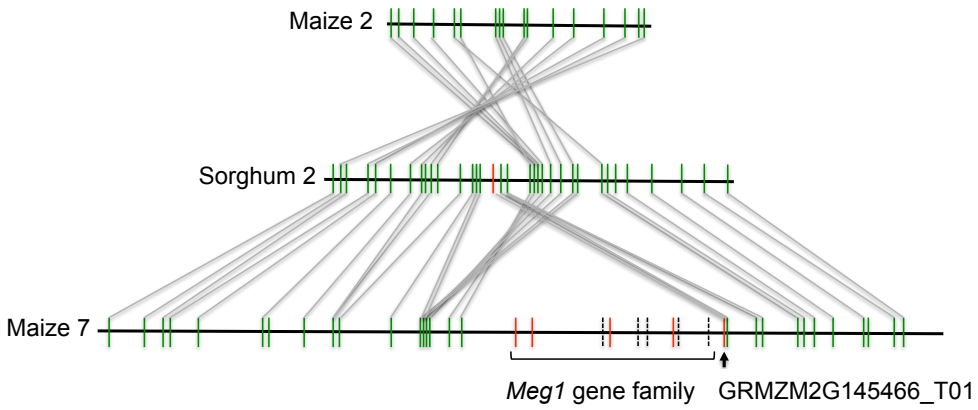

Figure 3.1: Schematic diagram comparing maize chromosomes 7 and 2, with sorghum chromosome 2. Many chromosomal rearrangements are indicated, though a high level of cluster conservation remains. The *Meg1* family can be found in maize 7, but has been lost from maize 2. GRMZM2G145466_T01 can be found at the end of the *Meg1* cluster and is the only gene conserved in sorghum 2.

In order to characterise this novel gene, a number of different experimental techniques were employed. A GUS reporter system was used to spatially and temporally assess the activity of the *ZmMlp1* promoter in developing seeds. This is a classical technique that allows visualisation of promoter active tissues by the development of a blue pigment in the site of activity (Jefferson *et al.* , 1987a). This technique is relatively simple to use, providing that stable transgenics can be generated, allowing the rapid processing of large numbers of samples in batch assays. Additionally, the analysis method is extremely simple, with visual screening achievable with the naked eye and imaging requiring only a light microscope. There are however, some disadvantages to this technique; the assay must be optimised in chemical composition and incubation time to avoid the pigment bleeding out into surrounding tissues. Furthermore, the sample is essentially fixed at the point of assay, a process that can cause tissue distortion. An alternative method

could be to use a GFP reporter which would allow live imaging of promoter activity at a much higher resolution. However, the sample analysis techniques required for GFP imaging are a lot more laborious, requiring extensive confocal or epifluorescent microscopy. Luciferase is another reporter system that could have been used, though this does not give the optical resolution required for this analysis.

Expression of the gene itself was assessed through a combination semi-qPCR, qRT-PCR and *in situ* hybridisation. *In situ* hybridisations have the advantage of spatially assessing the gene expression at the cellular level, though the technique is extremely labour intensive and requires a great deal of care in sampling and fixation to ensure mRNA integrity. Semi q-PCR is a dilution based PCR technique that yields gel bands of varying intensity according to target sequence concentration, though quantitative analysis of gels can be unreliable. qRT-PCR is considered the gold-standard in expression quantification, with unparalleled levels of sensitivity. However, great care must be taken with primer design, PCR condition optimisation and data analysis to ensure assay specificity and data reliability (Bustin *et al.* , 2009).

Protein localisation was carried out by both immunohistochemistry and GFP tagging approaches in order to gain the advantages and counter some of the disadvantages of both techniques. GFP labelling is an excellent technique for protein localisation studies, producing high quality images through confocal microscopy. However, this requires expensive and time consuming generation of stable transgenic lines, relatively laborious analysis methods and potential issues with steric alteration of the target protein causing localisation abnormalities. Antibody labelling however is relatively quick, simple, high-throughput and can yield reasonable quality images when combined with high quality sample preparation. However, due to the nature of antibody-epitope interaction, there remains the possibility for cross-reaction with other proteins containing sequence similarities. The application of this technique to a novel protein requires the generation of

a custom antibody and hence, acquisition of sufficient protein antigen for inoculation. To this aim, bacterial recombinant protein expression was employed, incorporating a His-tag for affinity purification. This system was chosen for its ease of use, high yield potential and low cost compared to the alternatives of yeast or plant based expression systems. However, the main disadvantage of using a bacterial expression system for a plant protein is the differences is protein folding machinery, leading to the potential for incorrectly folded product.

This chapter reports the combined use of these techniques to elucidate the expression characteristics of a novel gene, *ZmMlp1* and the biochemical characteristics of its product *Zm*MLP1 in the developing maize seed. This data is then considered in the context of the wider research area and used to drive functional based hypotheses and further research aims.

Results

3.2 Identification of MLP1 - a novel, small cysteine-rich protein

In silico analysis of the *ZmMlp1* DNA sequence revealed the general structure of the gene and promoter region (Figure 3.2 (a)). The coding region of the gene was found to be 578 bp in length, comprising of two exons separated by a centrally located 308 bp intron. After splicing, this gene produces a 270 bp mRNA sequence encoding the gene product, however previous work on EST analysis identified 5' and 3' untranslated regions (UTR), 83 and 105 bp in length respectively as indicated in Figure 3.2 (a). Interestingly, the promoter region was found to contain several TATA like elements and a 100 bp direct repeat sequence separated by 473 bp (Figure 3.2 (a)) which suggests that a sequence duplication event occurred during the evolution of the *ZmMlp1* promoter.

Analysis of the amino acid sequence (see 2.6.4) produced by the translation of *ZmMlp1*

yielded a small CRP with a sequence length of 89 residues, a calculated molecular weight of 9.6 kDa and a theoretical pI of 6.06. Sequence alignment of the *Zm*MLP1 protein with *Zm*MEG1 was performed (see 2.6.1) in order to assess the level of sequence similarity and conservation (Figure 3.2 (b)). The alignment shows a high degree of conservation at the amino acid level, with a sequence similarity score of 55.8 % which is further compounded by the similarity of the amino acid side chain properties as indicated by the colour scheme in Figure 3.2 (b). The main features conserved between the two protein sequences are; 8 highly spatially conserved cysteine residues (*Zm*MLP1 Cys_{49} Cys_{56}, Cys_{61}, Cys_{63}, Cys_{64}, Cys_{72}, Cys_{79}, Cys_{83} and *Zm*MEG1 Cys_{53}, Cys_{60}, Cys_{65}, Cys_{67}, Cys_{68}, Cys_{76}, Cys_{83}, Cys_{86} respectively) and two conserved tyrosine residues (*Zm*MLP1 Tyr_{62}, Tyr_{73} and *Zm*MEG1 Tyr_{66}, Tyr_{77} respectively). The sequence alignment shows significant stretches of conserved residues across the cysteine rich core and in the N-terminus, with non-conserved amino acids often conserving physiochemical properties (Figure 3.2 (b)).

Given the high degree of sequence similarity at the amino acid level, it could be hypothesised that *Zm*MLP1 and *Zm*MEG1 are functionally redundant genes, having arisen through gene duplication and subsequent sequence divergence. Alternatively, the sequence similarity could confer an evolutionarily stable substructure, upon which the sequence variation forms a functionally unique superstructure. Furthermore, the differences in the promoter regions of *ZmMlp1* and *ZmMeg1* could allow functional redundancy at the gene level, while expression control differs, leading to a divergent role for the *ZmMlp1* gene. In order to address these questions, I first analysed the expression pattern of the *ZmMlp1* gene.

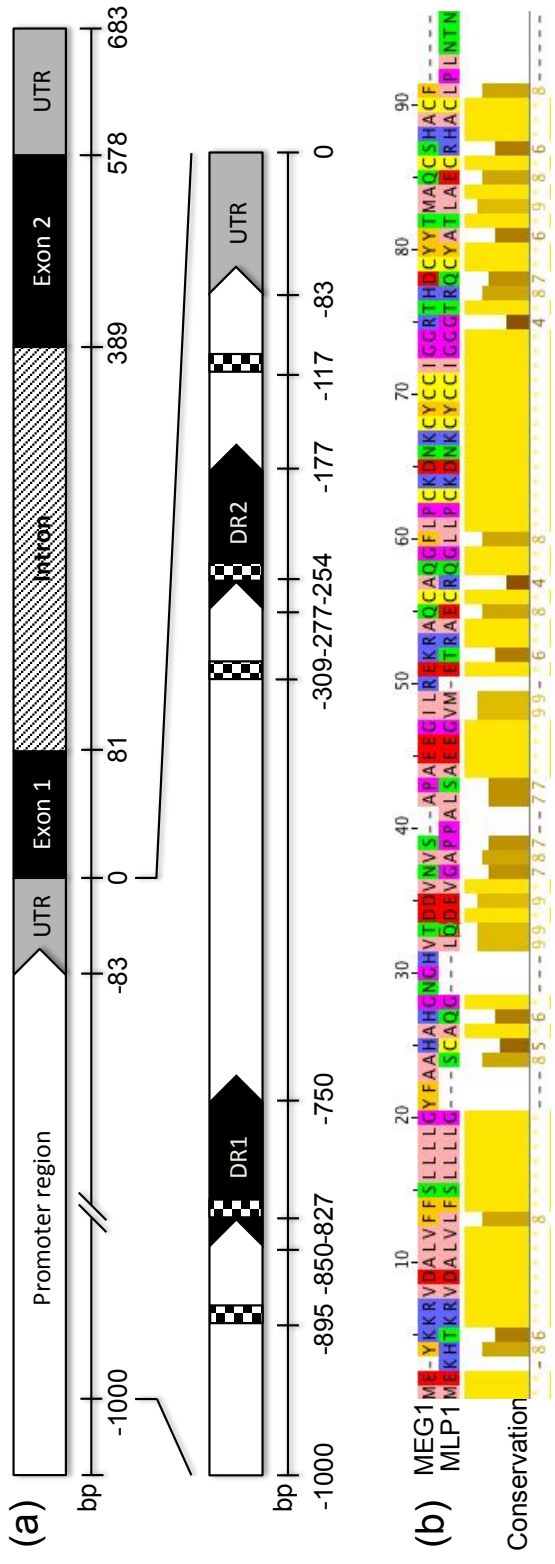

Figure 3.2: (a) Structure of the *Mlp1* gene indicating the promoter region (white), 5' and 3' untranslated regions (UTR) (grey), exons (black) and the single intron (hatching). Expanded promoter region illustrates a 100 bp direct repeat region (DR1 and DR2) and potential TATA box elements (chequered). (b) Amino acid sequence alignment of *Zm*MLP1 and *Zm*MEG1. Residues are colour coded according to physiochemical properties and sequence conservation indicated as a bar chart.

3.3 Expression analysis of *ZmMlp1*

Literature searches in the maize genome database using the gene accession number for *ZmMlp1* (GRMZM2G145466_T01) yielded expression microarray data (Figure 3.3)(Winter *et al.* , 2007; Sekhon *et al.* , 2011). From the eFP image, *ZmMlp1* transcripts were detected exclusively in the seed, at a developmental stage found between 6 and 12 DAP. This strict spatio-temporal expression pattern prompted analysis of the promoter activity to both verify the microarray data and further investigate the expression characteristics of *ZmMlp1*.

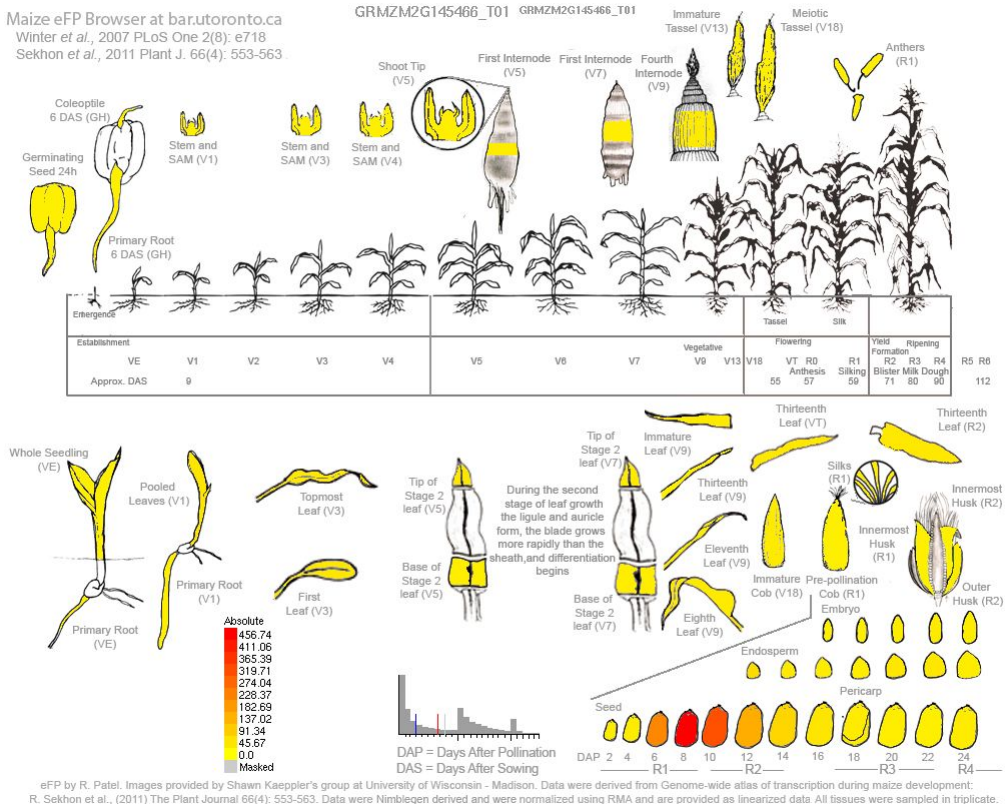

Figure 3.3: Electronic Fluorescent Pictograph (eFP) map of *ZmMlp1* (GRMZM2G145466_T01) expression generated from tissue specific microarray data. Colour of diagram indicates expression level of the *ZmMlp1* transcript as shown in the key. Expression of *ZmMlp1* is detected exclusively in the seed tissue between the developmental stages of 6 and 12 DAP. Adapted from (Winter *et al.* , 2007; Sekhon *et al.* , 2011).

3.3.1 The *ZmMlp1* promoter is spatially and temporally specific to the ESR in the early seed

In order to determine the activity of the *ZmMlp1* promoter (pro*Mlp1*), the β-glucuronidase (GUS) reporter system was used (Jefferson *et al.* , 1986, 1987a,b). The sequence of the promoter region containing both direct repeats (Figure 3.2 (A)) was extracted from the maize genome and used to generate a pro*Mlp1*:GUS fusion construct (Figure 7.1) which was then used to generate the expression plasmid pBIOS 1865 (Figure 7.6). pBIOS 1865 was then transformed into inbred wild-type maize of the A188 genetic background by Biogemma to generate approximately 60 independently transformed, transgenic plant lines of the construct line T01847. The independent lines selected for positive transformants containing only one transgene insertion (see 2.1.7). Initial screening of line T01847 was carried out by growing all independently transformed T0 plantlets (see 2.1.4) to maturity, back crossing with wild type pollen (see 2.1.5) and assaying the seeds for GUS expression at 8-10 days after pollination (DAP) (see 2.4.14). Seed stocks were then propagated by Biogemma through back-crossing with wild type pollen and the resulting seeds from ten GUS-expressing lines were transferred to Warwick for further analysis. Transgenic T01847 seeds from the ten lines with confirmed GUS expression were then grown to maturity (see 2.1.1) and crossed by sibling pollen (see 2.1.5). Ovules and developing seeds at 10 DAP were then assayed for GUS accumulation (see 2.4.14) to determine the spatial and temporal activity of pro*Mlp1*.

pro*Mlp1* was found to be highly active in both unfertilised ovules and developing seeds at 10 DAP (Figure 3.4). In the unfertilised ovules, GUS expression appears to be restricted exclusively to the embryo sac, with the signal concentrated in the central cell (Figure 3.4 (a-c)). Ten days after fertilisation, high level pro*Mlp1* activity was also observed, expressing GUS exclusively in the endosperm. Furthermore, the GUS signal in the endosperm appears to be localised to the tissue surrounding the basal part of the

embryo suspensor known as the embryo surrounding region (ESR) (Figure 3.4 (d-f)). This experiment has shown that pro*Mlp1* is specific to the central cell before fertilisation and the ESR after fertilisation. However, these data only illustrate the activity of the promoter and therefore it has limitations on the analysis of gene expression levels.

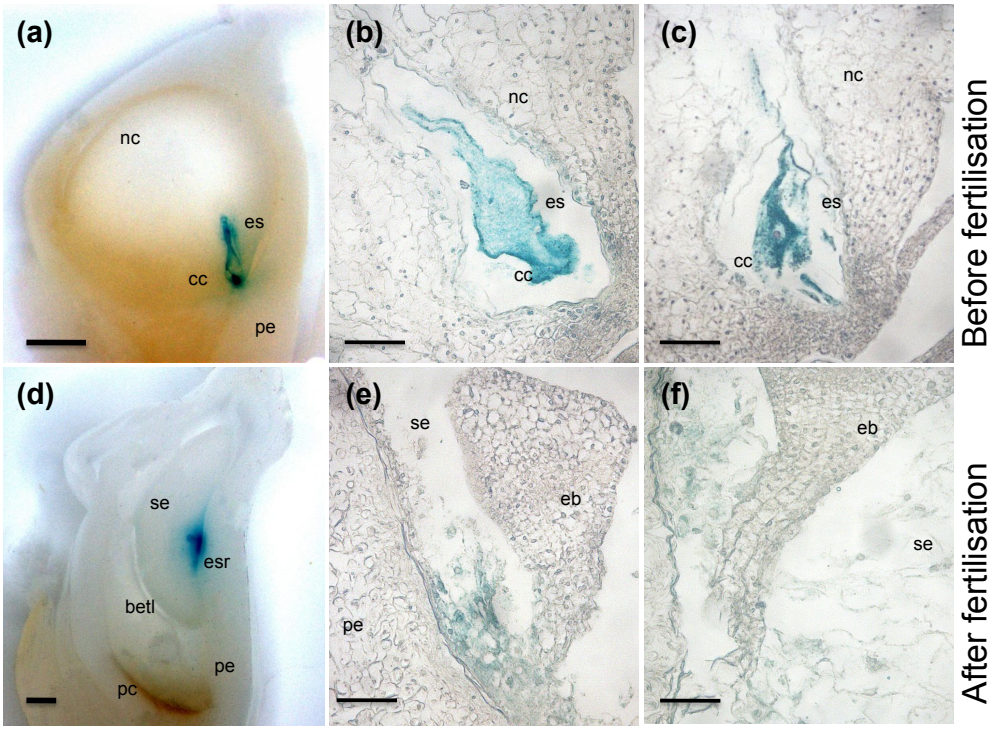

Figure 3.4: Spatial and temporal activity of the *ZmMlp1* promoter (pro*Mlp1*). Blue colour indicates GUS precipitate accumulation in pro*Mlp1*:GUS transgenic maize lines (T01847_054 sib crossed). (a-c) Ovules sampled before fertilisation, (d-f) developing kernels sampled at 10 DAP. (nc) nucellar material, (es) embryo sac, (cc) central cell, (pe) pericarp, (eb) embryo, (se) starchy endosperm, (esr) embryo surrounding region, (betl) basal endosperm transfer layer, (pc) placentochalazal region. (a and d) Whole fixed samples cut longitudinally. (b, c, e and f) formalin-fixed paraffin embedded (FFPE) samples sectioned longitudinally to 12 μm. Scale bars in (a and d) = 1 mm, (b, c, e and f) = 100 μm.

3.3.2 The *ZmMlp1* gene expression is temporally regulated

To establish the expression pattern of the *ZmMLP1* gene, wild type plants were grown to maturity (see 2.1.1) and crossed (see 2.1.5) for analysis. Unfertilised ovules and developing kernels were manually dissected from ears at 2, 4, 6, 8 and 10 DAP for RNA extraction (see 2.2.1), cDNA synthesis (see 2.2.3) and *ZmMLP1* expression quantification by both semi-qPCR and qRT-PCR (see 2.2.5 and 2.2.6) using oligonucleotides specific for *ZmMLP1* (see Table 7.2).

ZmMlp1 expression is detected in unfertilised ovules (Figure 3.5 (b)), complementing the pro*Mlp1*:GUS expression data seen in Figure 3.4 (b-d). After fertilisation, expression of the *ZmMlp1* gene increases from about 4 DAP peaking at around 8-10 DAP (Figure 3.5 (a and b)), consistent with the microarray data (Figure 3.3).

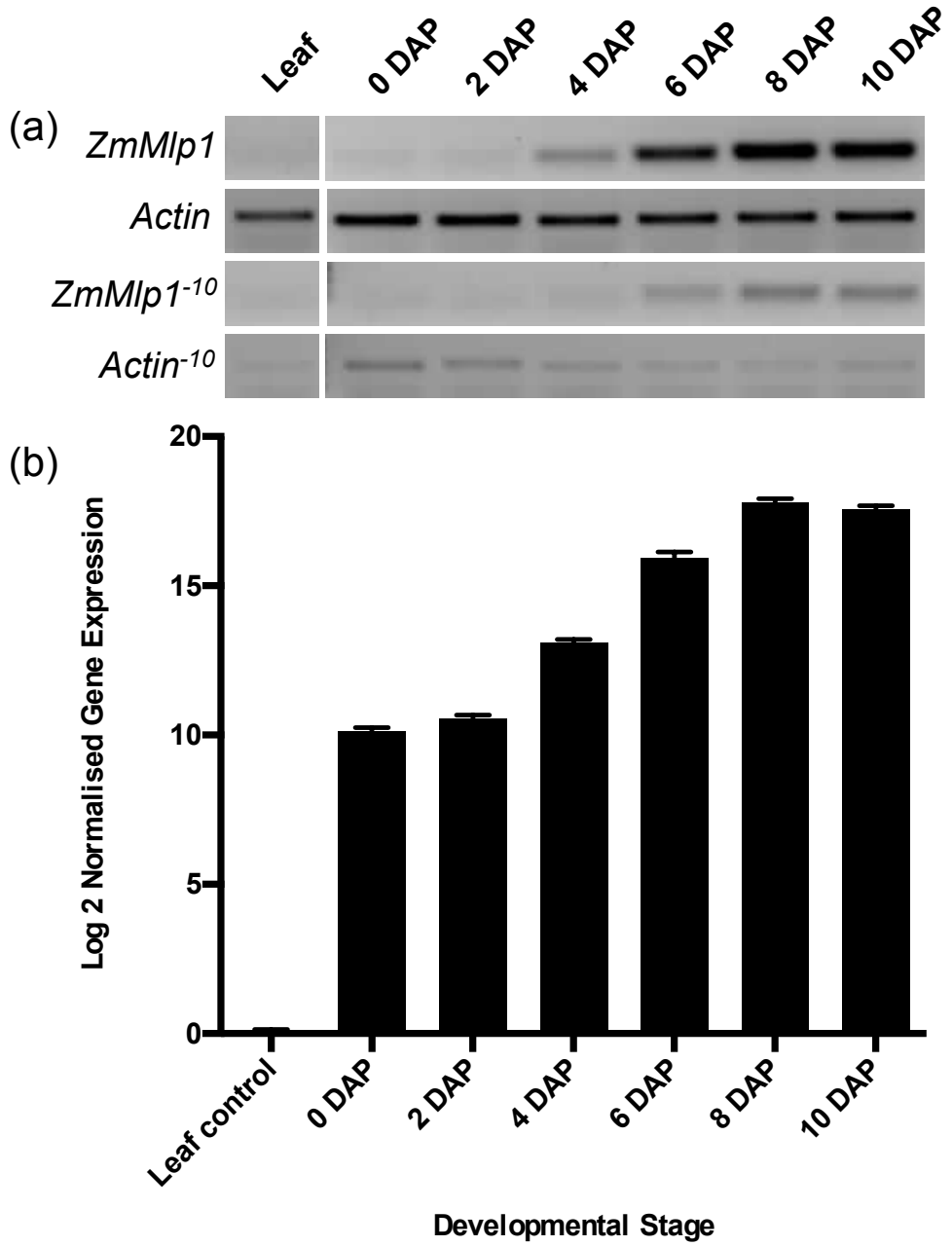

Figure 3.5: Temporal expression profile of the ZmMlp1 gene. (a) Semi-qRT-PCR *ZmMlp1* and *Actin* amplicons from 25 ng and 2.5 ng template cDNA respectively. (b) *ZmMlp1* expression analysis on whole seeds by qRT-PCR relative to the ubiquitous gene *ZmGapdh* (Gene identifier M18976). Error bars = Standard Error of the mean from technical triplicates.

To confirm that expression localisation at the mRNA level correlates well with the promoter activity and the temporal expression observations, *in situ* hybridisation experiments were performed (see 2.5.4). Probes were designed for the detection of the *ZmMlp1* mRNA transcript and used for *in situ* analyses of developing kernels at 8 and 10 DAP (Figure 3.6 (a-c) and (d-e)) respectively. *ZmMlp1* mRNA accumulation was observed in the ESR of the developing seeds (purple-brown precipitate in Figure 3.6 (b, c, e and f)). Efficacy of the mRNA probe was confirmed by control experiments using anti-sense probes which showed no detectable signal (Figure 3.6(a and d)).

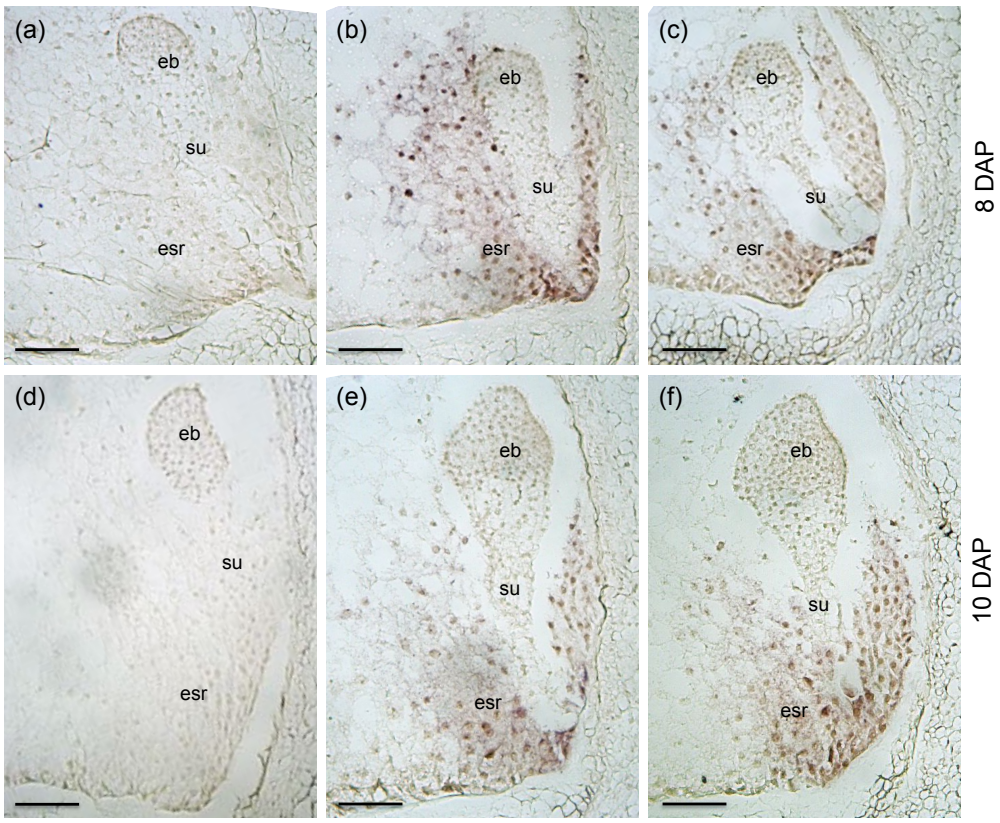

Figure 3.6: Analysis of *ZmMlp1* expression localisation. *In situ* hybridisation of *ZmMlp1* transcripts in fixed, developing maize kernels at (a-c) 8 DAP and (d-f) 10 DAP. (a and d) Sections analysed using probes designed using the *ZmMlp1* sense DNA strand, (b, c, e and f) probes designed using the anti-sense DNA strand. Purple-brown precipitate indicates presence of the *ZmMlp1* RNA probe. (esr) embryo surrounding region, (su) suspensor region, (eb) embryo. Scale bars = 100 µm.

Collectively, these data indicate that the *ZmMlp1* gene is spatially and temporally regulated, to restrict expression to central cells and ESR cells. However, further analysis is required to assess the precise expression of the *Zm*MLP1 protein.

3.4 MLP1 Protein Expression Analysis

Having reliably determined the gene expression pattern of *ZmMlp1*, the strict spatio-temporal regulation together with the previously discussed gene conservation evidence suggests that the gene product of *ZmMlp1* might be functionally important. Since there is a high degree of controversy over the correlation between mRNA expression levels and translated protein levels (Gygi *et al.* , 1999; Gry *et al.* , 2009; Greenbaum *et al.* , 2002), the next crucial step was to confirm the expression profile of the protein.

3.4.1 Expression and purification of MLP1

In order to investigate temporal and spatial *Zm*MLP1 expression, a reliable antibody was required. Due to the limited number of cells expressing *Zm*MLP1, isolation of sufficient quantity of the native protein for host animal inoculation was deemed impractical and therefore a recombinant protein expression approach was utilised.

Taking advantage of the short sequence length of *ZmMlp1*, a synthetic gene construct was designed (see 2.3.1) for expression of *Zm*MLP1 tagged with hexa-histidine in *E.coli* (Figure 7.2). Due to the presence of a predicted signal peptide, unlikely to be recognised in *E.coli,* a truncated sequence was used containing only the predicted mature length protein sequence. In order to achieve good expression yields, the sequence was optimised using the *E.coli* codon usage table before synthesis and cloning into a pET29a (Agilent Technologies) expression vector (see 2.3.2).

An inducible expression strain of *E.coli* was transformed with the *ZmMlp1* expression

vector and induced chemically (see 2.4.1). Protein was then extracted and the His-tag used to successively purify *Zm*MLP1 with nickel agarose beads under native and then denaturing conditions (see 2.4.8). The purified protein was then analysed by SDS-PAGE to determine the expression level and purification efficacy (Figure 3.7 (a and b)). Significant expression of *Zm*MLP1 was achieved with reasonable purification as shown by the strong, single band produced by the induced *Zm*MLP1 clone, which was not present in the induced empty vector control or the un-induced *Zm*MLP1 lanes in Figure 3.7 (a). The purified protein however had an apparent molecular weight of approximately 14 kDa, nearly double that of the expected 7.8 kDa of the *Zm*MLP1-His construct. This apparent shift in molecular weight could be caused by dimerisation of the protein which was too strong an interaction to break using the standard SDS-PAGE protocol or could be due to post translational modifications, occurring in the expression cells. More likely though, is that depending on the secondary structure adopted in the SDS micelle, the migration rate can be affected (Therien, 2001). Furthermore, the protein was isolated effectively under native conditions, indicating a high degree of solubility, suggesting that some degree of folding had taken place and that inclusion bodies were not formed.

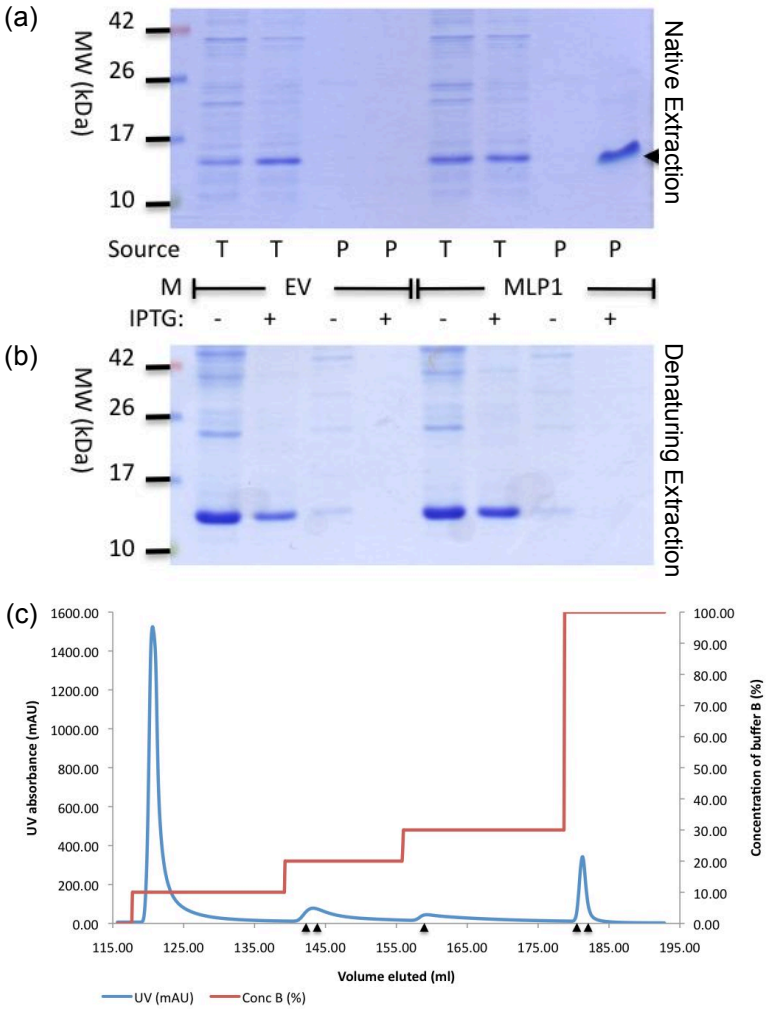

Figure 3.7: Expression and purification of *Zm*MLP1 from *E.coli.* (a and b) Coomassie stained SDS-PAGE gels of protein expressed from MLP1 and empty vector control (EV) clones purified under both native (a) and denaturing (b) conditions. Induction with IPTG indicated with -/+ while the source of the protein is indicated by T (total) and P (purified). Protein ladder lane indicated by 'M'. The arrow indicates expressed MLP1. (c) Chromatogram showing UV absorbance data of eluted protein from IMAC purification of MLP1 *E.coli* protein extract in conjunction with % eluting buffer (B). Arrows indicate fractions pooled.

In order to obtain sufficient *Zm*MLP1 protein for generation of an antibody, the expression system was scaled-up and optimised (see 2.4.1) before high-throughput purification was performed by IMAC (see 2.4.9). The chromatogram generated during the purific-

ation illustrates that four distinct peaks were eluted from the column at increasing concentrations of imidazole (Figure 3.7 (c)). The first peak represents protein that was bound non-specifically to the column, while the second, third and fourth peaks represent fractions containing *Zm*MLP1 as confirmed by SDS-PAGE analysis (not shown) (see 2.4.6). The differing affinities for the nickel column could be indicative of different structural isoforms of *Zm*MLP1, produced during incorrect folding in *E.coli.*

3.5 *Zm*MLP1 Antibody Production and Analysis

In order to confirm that the purified protein was in fact *Zm*MLP1, a purified sample was tryptically digested and analysed by mass spectrometry (see 2.4.10). The purified protein was positively identified as *Zm*MLP1, by the detection of 4 peptides with no contaminating *E.coli* proteins detected (Table 3.1). Purified *Zm*MLP1 was then sent to Eurogentec to generate a polyclonal antibody in rabbit (see 2.4.11).

The generated antisera was then tested by Western blot (Figure 3.8 (a)) using the antigen at decreasing concentrations in order to ascertain its efficacy at detecting *Zm*MLP1 (see 2.4.15). The antibody was found to be effective and so was used to detect *Zm*MLP1 in protein extracted from maize seeds. Developing wild type seeds were bisected as shown in Figure 3.8 (b) before protein was extracted from both the soluble and insoluble fractions and subjected to Western blot analysis (Figure 3.8 (b)) (see 2.4.3,2.4.15). A protein band was detected in the soluble fraction at slightly lower molecular weight than the *E.coli Zm*MLP1 positive control, which was enriched in the extract from the micropylar end of the seed. A band of the same size is also present in the lane corresponding to the top section of the seed. This could indicate that the band is not specific to the ESR, however, the technical difficulty in isolating the ESR in one half of the seed could have easily resulted in cross contamination with ESR proteins. The observed

Table 3.1: Peptide fragments identified by nanoLC ESI MS/MS after tryptic digestion of SDS-PAGE protein band. Scores and expected values calculated from Mascot protein identification software. * indicates a significant detection according to an expected value <0.05.

Peptide sequenced	**MLP1 position**	**Score**	**Expected**
LQDEVGAPPALSAEEGVMETRA	2 - 23	50	2.3e-005*
		113	1.1e-011*
		95	7.4e-10*
		73	8.8e-008*
		50	2.1e-005*
		36	0.00051*
		91	1.5e-009*
RQGLLPCKDNKC	26 - 37	18	0.03*
		28	0.0036*
		20	0.021*
KCYCCIGGGTRQ	36 - 47	45	3.3e-005*
		37	0.00021*
		42	5.8e-005*
RQCYATLAECRH	46 - 57	34	0.00039*
		72	5.9e-008*
		72	6.6e-008*
		58	1.6e-006*
		55	3.5e-006*

difference in mobility of the protein could be due to different structures adopted in the SDS-micelle (Therien, 2001) due to different folding or post-translational modifications.

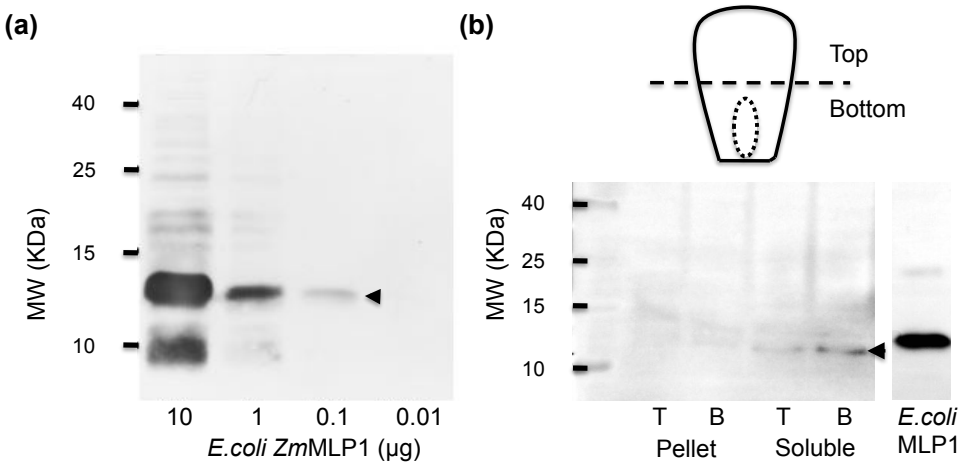

Figure 3.8: Evaluation of custom generated MLP1 antibody by Western blotting. (a) Serial dilution of MLP1 antigen performed and used in immunoblot to ascertain sensitivity of anti-MLP1 antisera. (b) 6 DAP maize seeds bisected as indicated and total protein extracted. Immunoblot carried out and band detected at approximately correct MW, present in bottom half of the seed, but not in the top (indicated by arrow).

3.6 *Zm*MLP1 is expressed in the central cell and ESR

Having obtained a useful antibody for *Zm*MLP1, the next step was to determine the protein expression pattern. In order to achieve this, an immunohistochemistry approach was taken, using the *Zm*MLP1 antibody to analyse the protein localisation in maize tissue sections. Wild type plants were grown to maturity (see 2.1.1), pollinated (see 2.1.5) and sampled at different developmental stages for fixation (see 2.5.1). Fixed samples were then embedded in paraffin wax and sectioned (see 2.5.2) before immunohistochemical analsysis (see 2.5.3). The detection method used alkaline phosphatase conjugated secondary antibodies and the synthetic substrate NBT/BCIP which produces a purple precipitate. Sections were first analysed using the pre-immune serum from the host antibody as a negative control to ensure there was no non-specific binding and no signal

was observed (not shown). In unfertilised ovules, protein localisation was detected in the embryo sac (Figure 3.9 (a)), though due to the tissue processing the central cell was observed to be collapsed meaning that protein accumulation in the central cell could not be completely confirmed.

At 2 DAP, very little signal was detected in the endosperm (Figure 3.9 (b)). This could be due to a lack of protein expression, or could be simply that due to the mostly free-nuclear structure of the endosperm at this developmental stage, there was not enough accumulated protein in any one area to reach the detection limits of the experiment. At 4 DAP however, clear and strong signal was detected in the ESR and appearing to decorate the surface of the protoderm of the embryo (Figure 3.9 (c and d)). There was also some localisation observed in the BETL (Figure 3.9 (d)), however, due to the similarity in amino acid sequence of *Zm*MLP1 and *Zm*MEG1, this is likely to be cross-reaction with *Zm*MEG1 in the transfer cells as their development initiates. At 6 DAP, the localisation of *Zm*MLP1 appears to remain in the ESR, which is beginning to diminish in size at this point (Figure 3.9 (e)). There is also less accumulation of *Zm*MLP1 on the embryo surface, with most of the signal concentrated around the suspensor region.

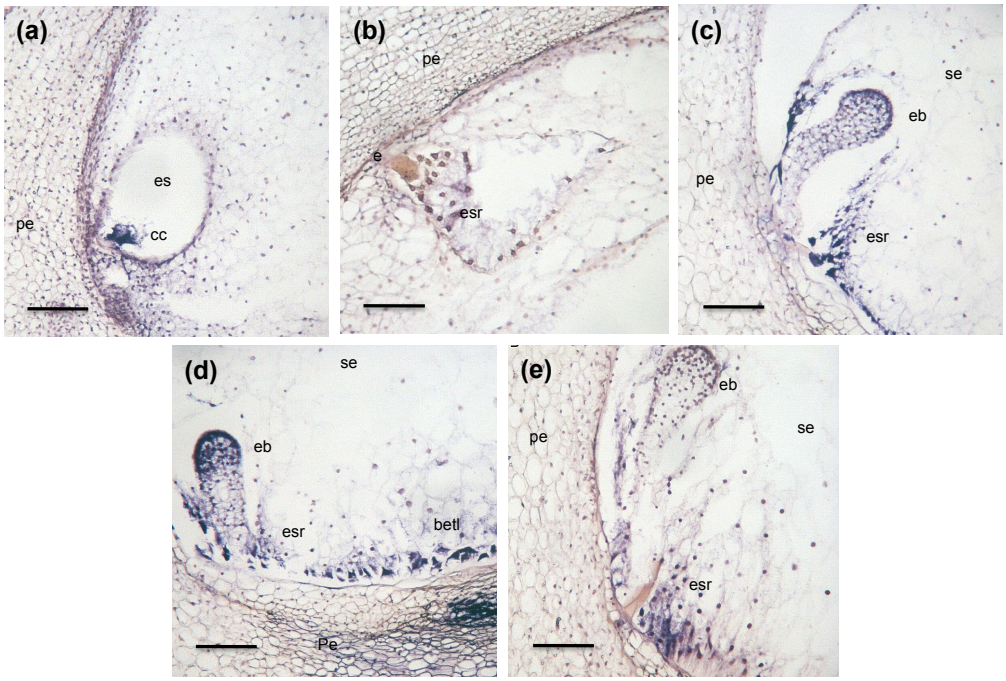

Figure 3.9: Cellular localisation of the MLP1 protein in FFPE maize seeds at different developmental time points by immunohistochemistry. All images present localisation of MLP1 on 12 µm WT seed sections using *E.coli* MLP1 generated primary antibody. Samples taken at (a) pre-fertilisation, (b) 2 DAP, (c-d) 4 DAP, (e) 6 DAP stages. MLP1 protein localisation indicated by deep purple precipitate. (es) embryo sac, (cc) central cell, (pe) pericarp, (eb) embryo, (se) starchy endosperm, (esr) embryo surrounding region, (betl) basal endosperm transfer layer. Scale bar = 100 µm.

To confirm the *Zm*MLP1 protein expression pattern observed using the antibody, a construct was designed to express *Zm*MLP1-GFP, driven by the native *ZmMlp1* promoter, pro*Mlp1* (Figure 7.3). Due to the predicted signal peptide cleavage site (see Figure 3.16 (a)), the GFP tag was inserted between the cleavage site and the predicted mature protein to prevent undesired cleavage of the tag, generating a mature *Zm*MLP1 N-terminal fusion. An N-terminal fusion was chosen over a C-terminal fusion due to the comparatively longer sequence of the N-terminal tail, a feature that would place the bulky GFP tag further from the cysteine rich core, reducing the chance of steric interference during protein folding. This construct was then used to generate the plas-

mid pBIOS3105 (see Figure 7.7) and transformed into A188 maize by Biogemma to generate the *Zm*MLP1-GFP transgenic construct line T02492 (see 2.1.8).

T02429 plantlets were generated from independently transformed calli, yielding 25 lines which were transferred to Warwick, grown to maturity (see 2.1.4) and pollen back-crossed into wild type plants (see 2.1.5). Segregating A188 x T02492 seeds were then germinated, genotyped for BAR (see 2.2.10 and Table 7.1) and grown to maturity (see 2.1.1). Sib-crosses were then performed and the developing kernels sampled at different time points for analysis by confocal and epifluorescent microscopy (see2.5.5 and 2.5.6). Unfertilised ovules were analysed and GFP signal was detected in the central cell of the embryo sac in multiple, independently transformed sub-lines (Figure 3.10) and was confirmed to be genuine signal by analysis of GFP negative ovules (Figure 3.10 (a and d)). Central cell identity was confirmed morphologically by the large and highly vacuolated nature of the GFP containing cells (Figure 3.10 (g and h)), features typical of the maize central cell (Kranz *et al.* , 1998).

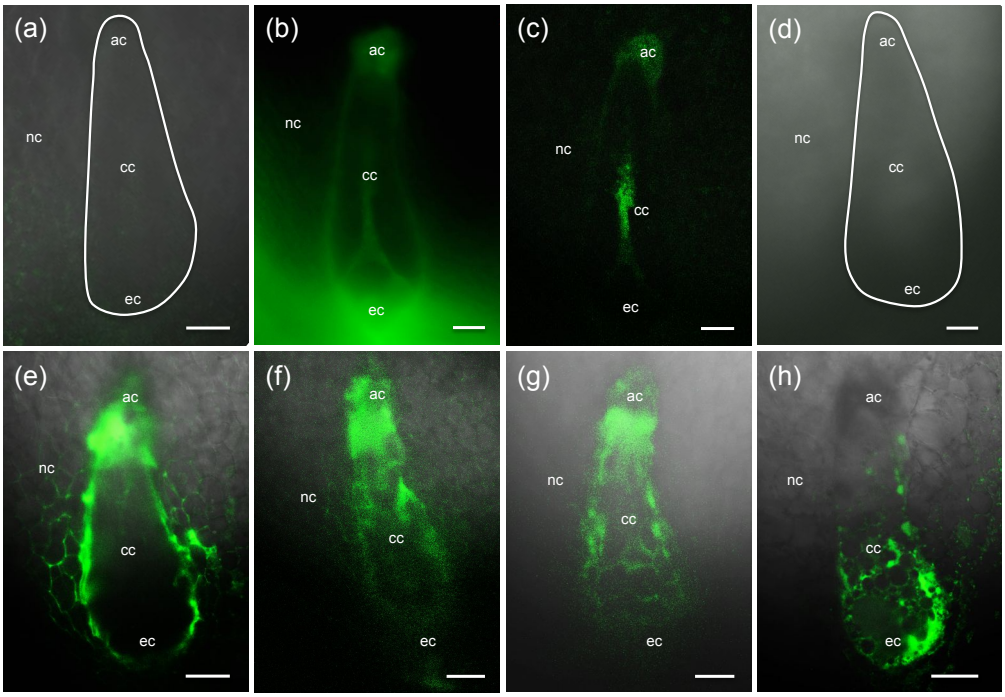

Figure 3.10: Microscopic images showing MLP1-GFP expressed in unfertilised maize ovules from the line T02492. (a and d) Line T02492 CM-07639C-3 and CM-07640H-3 GFP-negative ovules respectively. (b-c and e-h) GFP expression detected in the central cell of the embryo sac in lines T02492 CM-07639C-3 and CM-07640H-3 respectively. (b, c, f and h) Intact embryo sacs. (e) Embryo sac sliced longitudinally. (a, c, e, f, g and h) Single optical sections obtained by laser scanning confocal microscopy. (b and d) Epifluorescence UV microscopy. cc = central cell, ac = antipodal cells, nc = nucellar tissue, ec = egg cell. Scale bars = 50 μm

At 4 DAP, GFP signal was observed in the ESR which at this developmental stage was completely enveloping the proembryo (Figure 3.11 (a-c)). Signal also appeared to be localising to the surface of the proembryo (Figure 3.11 (d)), consistent with the localisation data obtained from the immunohistochemistry experiments. 6 DAP seeds exhibited much the same expression pattern with GFP signal visible in the ESR and on the surface of the embryo, which begins to emerge from the ESR at this stage (Figure 3.12 (a-d).

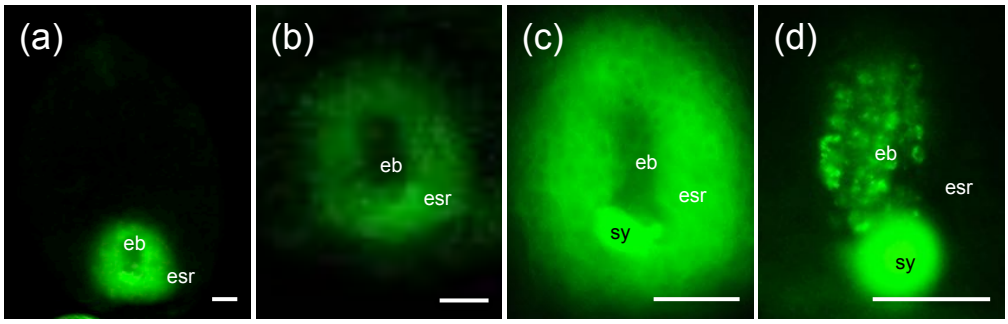

Figure 3.11: Epifluorescent UV microscopic images showing MLP1-GFP expressed in developing maize seeds form line T02492 CM-07640H-3 at 4 DAP. (a-c) GFP signal detected in the ESR and (d) on the surface of the developing pro-embryo. Bright signal also observed in the degenerated synergid (c and d) which was verified to be autofluorescence by analysis of GFP negative seeds. ESR = embryo surrounding region, eb = embryo, sy = degenerated synergid. Scale bar = 100 µm

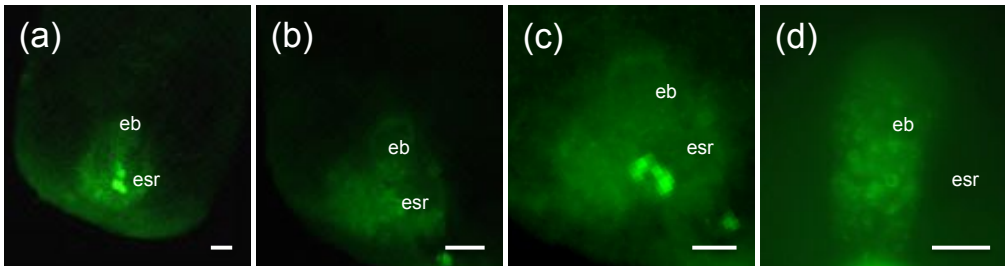

Figure 3.12: Epifluorescent UV microscopic images showing MLP1-GFP expressed in developing maize seeds form line T02492 CM-07640B-2 at 6 DAP. (a-c) GFP signal detected in the ESR and (a-d) on the surface of the developing pro-embryo as it emerges from the ESR. Autofluorescence again observed in the degenerated synergid (a and c). esr = embryo surrounding region, eb = embryo. Scale bar = 100 µm

At 8 DAP, strong GFP localisation to the ESR was observed (Figure 3.13 (a-g) and data was obtained that clearly demonstrates GFP accumulating on the surface of the embryo, on both the suspensor and protoderm, but not expressing in the cells of the embryo itself (3.13 (d and f)). In high magnification images the GFP signal was observed to be accumulated at the outer extremes of the cells of the ESR, suggesting that GFP is being secreted from these cells (Figure 3.13 (g)). At 10 DAP, seeds were shown to exhibit GFP localisation in the degenerated ESR, surrounding only the cells of the suspensor

(Figure 3.14 (b, c and e)). There was also evidence of GFP decorating the cells of the basal suspensor as imaged in Figure 3.14 (d) and (f).

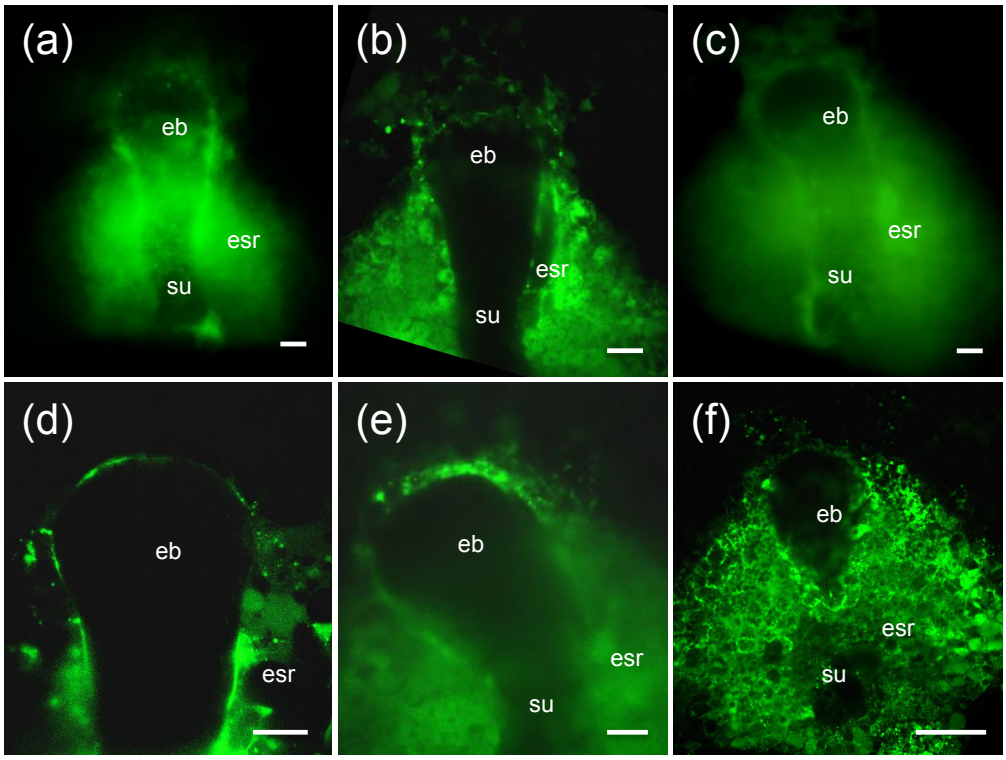

Figure 3.13: Laser scanning confocal microscope single optical sections showing MLP1-GFP expressed in developing maize seeds form line T02492 CM-07640H-3 at 8 DAP. (a-c) GFP signal detected in the ESR, surrounding the suspensor region of the embryo. (d-f) GFP signal localised to the protoderm of the embryo while no signal is observed in the cells of the embryo itself. (g) GFP signal localised to the outer extremes of the ESR cells. esr = embryo surrounding region, eb = embryo, su = suspensor region. Scale bar = 50 μm

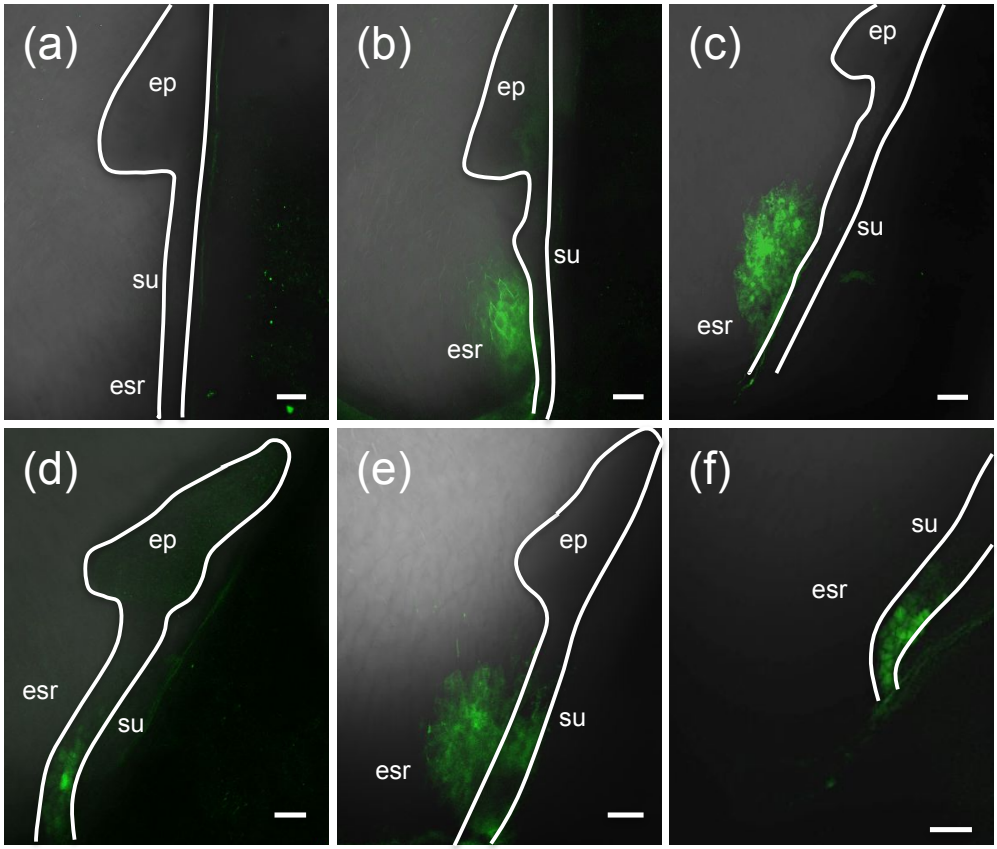

Figure 3.14: Laser scanning confocal microscopy images of developing maize seeds at 10 DAP. (a) Line T02492 CM-07640B-2 GFP-negative seed. (d and f) GFP expression detected in the diminished ESR and decorating the surface of the embryo suspensor from line G22. (b, c and e) GFP signal detected in the ESR in lines T02492 CM-07640G-3, CM-07640G-3 and CM-07640H-3 respectively. ESR = embryo surrounding region, ep = embryo proper, su = suspensor region. Scale bars = 100 μm.

Despite the precautions taken in the design of the construct, the possibility remained that the GFP could be cleaved from *Zm*MLP1 *in planta*, generating spurious GFP signal. In order to verify that the GFP localisation imaged in Figures 3.10-3.14 was in fact *Zm*MLP1-GFP and not a cleavage product, protein was extracted from *Zm*MLP1-GFP seeds at 8 DAP and analysed by Western blot (see 2.4.17 and 2.4.15). Immunodetection with an anti-GFP antibody of protein extracted from *Zm*MLP1-GFP endosperms yielded a band approximately 10-15 kDa higher in molecular weight than the GFP pos-

itive control (Figure 3.15 (a-c)). Duplicate analysis was carried out using the *Zm*MLP1 antibody which was able to detect the same band (Figure 3.15 (d)) but was not shown to detect the GFP positive control (Figure 3.15 (f)). These analyses indicate that the GFP signal imaged by confocal microscopy was indeed still fused to the *Zm*MLP1 protein and together with the immunohistochemistry data, present a reliable localisation study of *Zm*MLP1.

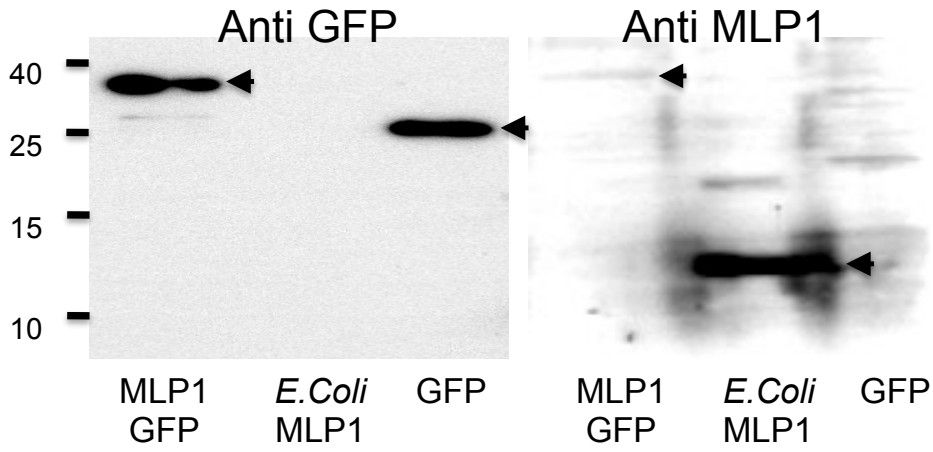

Figure 3.15: Detection of *Zm*MLP1-GFP from isolated endosperms at 8 DAP by Western blot. (MLP1-GFP) Protein extracted from maize *Zm*MLP1-GFP line G34 endosperms. (*E.coli* MLP1) Recombinant *Zm*MLP1 protein positive control. (GFP) GFP protein positive control. (Anti-GFP) Detection carried out using anti-GFP antibody (Miltenyi Biotec). (Anti-MLP1) Detection carried out using anti-MLP1 antibody. Arrows indicate: *Zm*MLP1-GFP, GFP, and Recombinant *Zm*MLP1.

3.7 *Zm*MLP1 contains a secretory signal peptide

As previously mentioned, *In silico* analysis of the *Zm*MLP1 amino acid sequence identified a predicted cleavage site for a signal peptide. This signal was identified using SignalP 4.0, an online tool capable of detecting cleavage sites in amino acid sequences and distinguishing the cleaved sequence from transmembrane helices based on their hydrophobicity characteristics(Petersen *et al.* , 2011). Peptide cleavage was predicted

between G_{25} and L_{26}(Figure 3.16 (a)). To determine if this signal peptide was able to facilitate protein secretion, sub-cellular localisation experiments were carried out using *Zm*MLP1-GFP in *Nicotiana benthamiana* leaves. This system was chosen as it is a well characterised and robust method for localisation studies of tagged proteins (Fischer *et al.* , 1999).

Agrobacterium tumefaciens mediated transient expression was used to express a *Zm*MLP1-GFP construct under the control of a CaMV 35S promoter (Figure 3.16 (b)) in leaf epidermis of *Nicotiana benthamiana* (see 2.1.2, 2.1.10 and 2.1.11). Two GFP control constructs (GFP-HDEL and GFP-LTI6b) with characterised expression patterns were also expressed in leaves to use as a comparison. GFP-HDEL is an endoplasmic reticulum (ER) targeted GFP construct (Batoko *et al.* , 2000) (Figure 3.16 (c)) while GFP-LTI6b (low temperature induced protein, accession AAC97512) is a construct consisting of a 6 kDa peptide fused to GFP, known to localise to the plasma membrane (PM) (Cutler *et al.* , 2000; Kurup *et al.* , 2005) (Figure 3.16 (d)). Two days post infection, leaves were sampled for expression analysis by laser scanning confocal microscopy (see 2.5.5). Upon imaging, the GFP signal from all three constructs was found to localise to the outer extremes of the cell, with ER localisation observed in the GFP-HDEL samples (not shown). Plasmolysis experiments were performed to ascertain if GFP signal was accumulating in the apoplastic space of the leaf cells, or perhaps binding to the membrane (see 2.5.5), however, GFP was not observed in the apoplast. Difficulty in imaging GFP in the apoplast may have been due to protein instability, or more likely, the low pH conditions of the apoplast diminishing the GFP fluorescence (Tsien, 1998; Grignon & Sentenac, 1991), an issue which could be solved by using a less pH sensitive fluorophore such as RFP or mCherry.

To establish whether *Zm*MLP1-GFP was in fact present in the apoplast and therefore confirm secretion, the apoplast was extracted from transformed *Nicotiana benthamiana*

leaves and the protein analysed by Western blot (see 2.4.16 and 2.4.15). Apoplastic fluid from WT leaves and those expressing *Zm*MLP1-GFP and GFP-HDEL as confirmed by confocal analysis was analysed using a GFP antibody (3.16 (e)). The apoplast from both the WT and GFP-HDEL controls were shown to contain no detectable GFP as expected, due to retention of GFP in the ER (Figure 3.16 (e)). This result also indicates that the cells were not ruptured during the apoplastic fluid isolation. Apoplast from the *Zm*MLP1-GFP leaves however, contained GFP as detected by the antibody (3.16 (e)). The protein was then extracted from the apoplast-depleted leaves and subjected to the same analysis, demonstrating that GFP-HDEL was present and detectable, and that no significant amount of GFP was retained in the cells of the *Zm*MLP1-GFP leaves (Figure 3.16 (e)). Interestingly though, the band detected from the *Zm*MLP1-GFP leaf apoplast, was of the same molecular weight as GFP, indicating that *Zm*MLP1 had been cleaved from the GFP tag. This could be due to the *Zm*MLP1 protein being expressed in the wrong organism and indeed, the wrong tissue type. If the *Zm*MLP1 protein presented an issue of toxicity to the leaf, it may have been targeted for cleavage and degradation. Collectively though, these data support the hypothesis that *Zm*MLP1 contains a signal peptide which is able to target protein for secretion and offers further evidence that *Zm*MLP1 is secreted from expressing cells in maize.

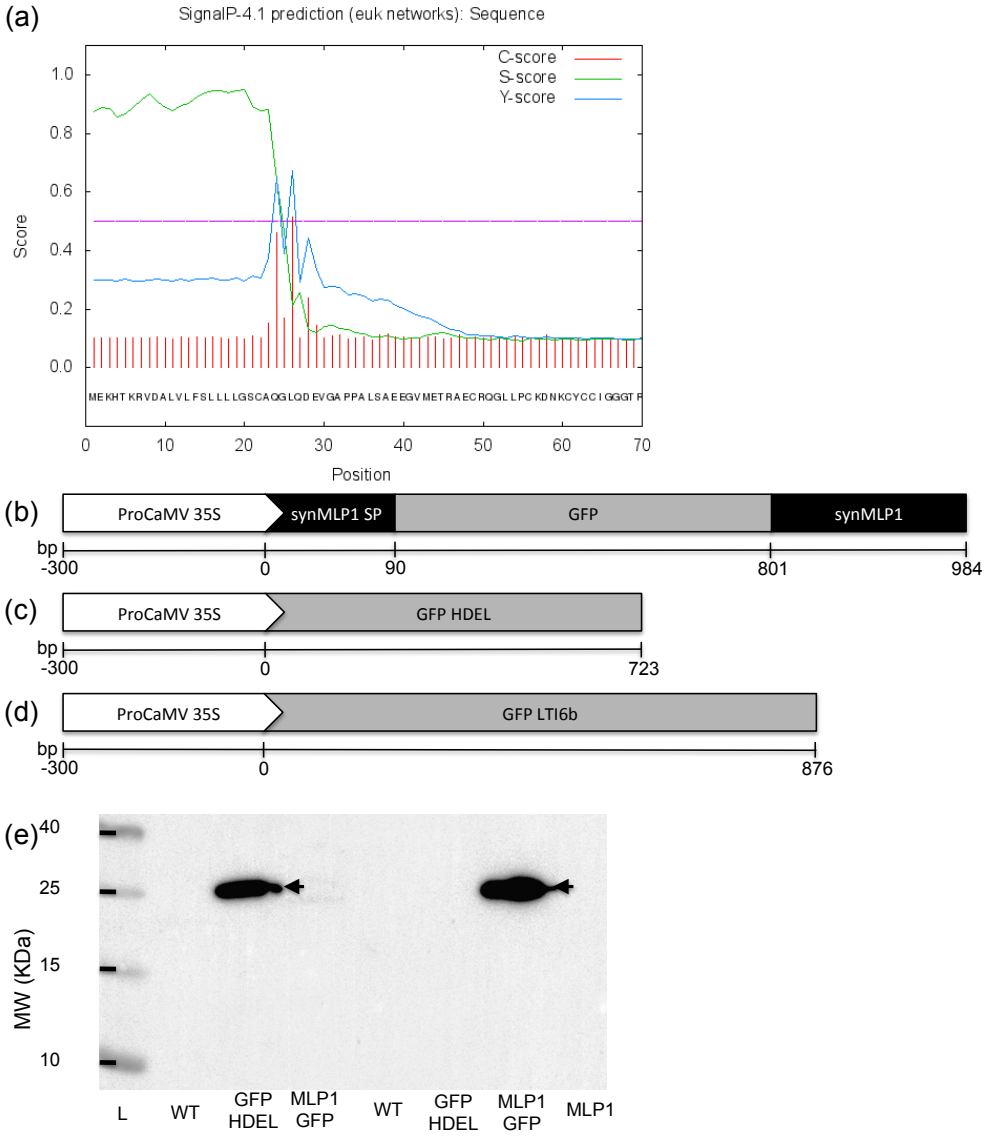

Figure 3.16: (a) *In silico* analysis of *Zm*MLP1 signal peptide cleavage site as predicted by the SignalP 4.1 online server (Petersen *et al.* , 2011). Peptide cleavage predicted between G_{25} and L_{26}. (b) Synthetic MLP1-GFP expression construct under the control of a 35S constitutive promoter. (c) 35S expression construct for GFP-HDEL. (d) 35S expression construct for GFP-LTI6b. (e) Western blot analysis using an anti-GFP antibody showing: (WT) wild type protein extract control. (GFP-HDEL) Protein extracted from leaves expressing GFP-HDEL. (MLP1-GFP) Protein extracted from leaves expressing MLP1-GFP. (Total protein) Protein extracted from apoplast depleted leaves. (Apoplastic protein) Protein extracted from apoplastic fluid.

3.8 Discussion

The maize ESR has been characterised morphologically (Opsahl-Ferstad *et al.* , 1997) and to some extent, by expression analysis of ESR specific genes such as *ZmEsr1-3* (Opsahl-Ferstad *et al.* , 1997; Bonello *et al.* , 2000, 2002), *ZmESR-6* (Balandín *et al.* , 2005) and *ZmAE1* and *ZmAE3 (Magnard* et al. *, 2000).* The function of the ESR however, remains largely hypothetical with only sporadic evidence for roles in pathogen defence (Balandín *et al.* , 2005) and embryo-endosperm interaction (Schel *et al.* , 1984; Opsahl-Ferstad *et al.* , 1997). Extensive analysis of a homologue of the BETL specific gene, *ZmMeg1* has identified a novel ESR and central cell specific gene, *ZmMlp1*, with a tightly controlled expression pattern which may help to further the understanding of the role of the ESR in maize seed development.

The conservation pattern of *ZmMlp1* both within and across species not only suggests an important functional role, but also implicates *ZmMlp1* as the ancestral gene that gave rise to the *ZmMeg1* family during maize domestication (Xiong *et al.* , 2014). Interestingly, the resolved expression pattern bears very little resemblance to that of *ZmMeg1* which is expressed exclusively in the transfer cells (Gutierrez-Marcos *et al.* , 2004; Costa *et al.* , 2012). In contrast, both the *ZmMlp1* gene and *Zm*MLP1 protein are expressed in the central cell and the ESR after both before and after fertilisation respectively, with expression levels rising progressively until 8-10 DAP. This increasing expression pattern presents its own challenges in collecting equivalent, comparable samples. Although the maize plants were grown in controlled environment glasshouses, the weather can still have a significant effect on developmental rates, with hotter and sunnier days accelerating the growth and development of the seeds. To control for this variability, control plants were grown at the same time, in the same glasshouse compartment ensuring consistency. However, this does not control for variation between batches, leads to an inconsistency between DAP and developmental stage which had to be resolved

by sampling, dissection and subjective observation to gauge developmental progression. This expression pattern bears some similarity to the ESR genes *ZmEsr1-3*, *ZmESR-6*, *ZmAE1* and *ZmAE3* (Opsahl-Ferstad *et al.* , 1997; Bonello *et al.* , 2000, 2002; Balandín *et al.* , 2005; Magnard *et al.* , 2000), though none of these genes has demonstrated central cell expression. The expression of *ZmMlp1* in fact, closely resembles that of *ZmEBE* which is expressed in the central cell, ESR and BETL (Magnard *et al.* , 2003), though *ZmMlp1* shows no BETL expression. Thus, *ZmMlp1* has been shown to exhibit a novel expression pattern in maize, localising exclusively to the central cell and ESR.

Preliminary investigation into the sequence of pro*Mlp1* identified a large direct repeat sequence, indicating a sequence duplication event. This could be evidence of an evolutionary protective strategy, safeguarding the expression of *ZmMlp1* through promoter element redundancy, or it could confer expression enhancement, either of which indicates selective pressure on the pro*Mlp1* promoter. Given that there are now several genes manifesting ESR specific expression in maize, it would be interesting to investigate the promoters of these genes more extensively to try and identify any common regulatory elements responsible for these expression characteristics. It would be useful from a biotechnology point of view to identify the specific elements that confer the different tissue specificities, to allow greater control over expression of useful transgenes in these cell-types.

The *Zm*MLP1 protein was localised to the central cell and ESR using two different approaches: immunohistochemistry and GFP fusion. This dual strategy was taken in order to overcome the disadvantages of both techniques for this application. Due to the high degree of similarity in amino acid sequence between *Zm*MLP1 and *Zm*MEG1 which is expressed in the endosperm at a similar developmental stage, there was the possibility of cross-reaction by the generated poly-clonal antibody. There is also the possibility of cross-reacting with unknown proteins containing similar epitopes and loss

of signal due to inadequate protein fixation. Additionally, the epitopes accessible in the recombinant protein may be internalised in the native protein due to alternative protein folding which could have a deleterious effect on signal strength. There was some apparent signal observed in the BETL region which is probably attributable to cross reaction with the similar *Zm*MEG1 protein. This could call into question the validity of the observed signal in the ESR, however, no other MEG family proteins have been observed in the ESR, MEG proteins are not expressed at the earlier developmental stages (Xiong *et al.* , 2014) and this expression pattern aligns well with the observed promoter activity.

In order to negate the aforementioned issues, the *Zm*MLP1-GFP fusion experiments were carried out as a second line of investigation. While the GFP system is well characterised, for this application it would not have been sufficient for protein localisation as a stand-alone experiment due to the small molecular size of *Zm*MLP1 compared to the relatively large GFP tag. Adding a tag, in this case, three times the size of *Zm*MLP1, could have adverse effects on protein expression, stability, targeting and function. The similarity of the results observed using both localisation approaches however, supports the conclusion that *Zm*MLP1 is expressed in the central cell and ESR with a high degree of confidence. Additional experimental work to confirm this further could be carried out by using an antibody for *Zm*MEG1 to confirm no ESR localisation, however after variety of approaches, a recombinant version of *Zm*MEG1 could not be expressed in a soluble form, making antibody generation more challenging.

The GFP fusion experiments allowed high definition imaging of *Zm*MLP1, showing what appeared to be decoration of the protoderm and suspensor of the embryo but no expression from the embryo cells. However, the surface signal could have been due to ESR tissue containing the fluorophore clinging to the surface of the embryo. Further experiments need to be conducted to determine the cell identity at the interface between

the embryo and the ESR, perhaps using a different fluorophore to label the ESR using one of the reported ESR specific genes such as *ZmESR6.* This however, would be a costly and time consuming approach, requiring the generation of new transgenic lines and the successful introgression of the two constructs into one line. Another approach might be to label ESR proteins with a fluorescent antibody such as FITC and perform dual imaging. However, *Zm*MLP1 observed on the embryo protoderm combined with *in situ* mRNA hybridisation showing no mRNA present on the embryo surface suggests that *Zm*MLP1 accumulates in the cells of the ESR, is secreted and most likely interacts with an as yet unidentified co-factor on the embryo.

The observed protein localisation prompted the investigation into the secretory properties of *Zm*MLP1, through transient expression in tobacco, which confirmed the hypothesis that the predicted *Zm*MLP1 signal peptide is sufficient for apoplastic secretion. This secretion is consistent with the behaviour of the ESR proteins ESR1-3 (Bonello *et al.* , 2002) and ESR6 (Balandín *et al.* , 2005) together with the hypothesised function of the ESR as a tissue facilitating interaction between the endosperm and embryo during embryogenesis (Schel *et al.* , 1984; Opsahl-Ferstad *et al.* , 1997). This approach however, was not ideal as a different host plant was used and could have been the cause behind the observed instability of the GFP tag, which was apparently cleaved from the *Zm*MLP1 protein. A better approach might have been to express the protein in maize leaves for imaging, though this is a less well developed technique as maize leaves are a more recalcitrant tissue type for apoplastic transformation and extraction. Furthermore, given the highly restricted expression pattern for *ZmMlp1* shown in this work, it might be that *Zm*MLP1 has cytotoxic properties in the wrong cell type, or that the protein would be unstable.

These observations offer a putative role for *Zm*MLP1 as a signalling component potentially similar to the related embryo surrounding factor (ESF) proteins studied in

Arabidopsis which govern embryo pattern formation during seed development (Costa *et al.* , 2014). Alternatively, *Zm*MLP1 could be acting as an endosperm regulatory factor, affecting the balance in development of embryo and endosperm size as observed in the case of *Zm*INVINH1 (Bate *et al.* , 2004) in maize and *GIANT EMBRYO* in rice (Yang *et al.* , 2013; Nagasawa *et al.* , 2013). As previously discussed, *Zm*INVINH1 is an invertase inhibitor CRP with expression restricted specifically to the apoplast of the ESR, indicating a role in the regulation of endosperm/embryo development rates by shielding the embryo from the high glucose concentrations found in the endosperm (Bate *et al.* , 2004). *GIANT EMBRYO* (*GE*) encodes the cytochrome P450 protein CYP78A13, which is expressed in the interfacing tissue between the rice embryo and endosperm, much like the ESR of maize. CYP78A13 acts to negatively regulate the size of the rice embryo during its development as shown by the *GE* phenotype in loss-of-function mutants (Nagasawa *et al.* , 2013).

4 Functional Characterisation of *ZmMlp1*

4.1 Introduction

Plant CRPs are a diverse protein class with a wide range of functions including pathogen defence, bacterial symbiosis, development, pollen recognition, pollen guidance and seed function across a variety of plant species (Marshall *et al.* , 2011). In maize, the imprinted CRP gene *Maternally Expressed Gene1* (*ZmMeg1*) is expressed exclusively in the Basal Endosperm Transfer Layer (BETL) of the developing kernel (Gutierrez-Marcos *et al.* , 2004). The gene product, *Zm*MEG1 is a small CRP involved in sugar partitioning within the seed through regulation of BETL transfer cell development (Costa *et al.* , 2012). Due to the high degree of similarity at the amino acid sequence level, it was originally hypothesised that *Zm*MLP1 could have a similar function to *Zm*MEG1 or even functional redundancy. However, the characterisation of *Zm*MLP1 described in chapter 3 revealed a unique and highly regulated expression pattern, restricted to central cells and the ESR.

Central cell and ESR expression suggests perhaps a different role for *Zm*MLP1 as the ESR has been implicated in communication between the embryo and endosperm (Schel *et al.* , 1984; Opsahl-Ferstad *et al.* , 1997). One hypothesis might be that the ESR functions, in part, to regulate the translocation of nutrients from the endosperm into the developing embryo. Therefore, *Zm*MLP1 could function similarly to *Zm*MEG1 in the BETL, having a regulatory role in the development of the ESR cells. Alternatively, given the expression characteristics of the *Zm*MLP1 protein discussed in chapter 3, particularly the secretion of *Zm*MLP1 and its detection on the surface of the developing embryo, *Zm*MLP1 could be acting as a signalling component for embryo-endosperm cross-talk.

Having characterised the expression pattern of *ZmMlp1*, the aim of this chapter is to investigate its functional roles and properties and contextualise these with considera-

tion of other MLP1-like homologues and other members of the CRP family. In order to assess the function of a gene of interest, two common strategies are to investigate phenotypic differences between wild type and loss-of-function mutants, or conversely, over-expression lines. To this aim, mutant database (UniformMu, Maize GDB (McCarty *et al.* , 2005)) searches were performed to try and identify a mutant maize line with a mutation mapped to the locus of *ZmMlp1*, however, no mutants were identified. Since mutagenesis experiments are generally non-specific and extremely slow, a strategy was developed to generate transgenic maize plants with depleted or augmented *ZmMlp1* expression.

One approach that was considered to elucidate the function of *ZmMlp1* was to generate over-expression transgenics, either expressing *ZmMlp1* using tandem repeats with the native promoter, or to express *ZmMlp1* under the control of an alternative promoter causing mis-expression. However, transgene over-expression in plants has been shown on many occasions to initiate gene silencing mechanisms, leading to post-transcriptional gene regulation by small RNAs (Flavell, 1994; Baulcombe, 1996; Finnegan *et al.* , 2001; Vaucheret & Fagard, 2001; Matzke *et al.* , 2002; Wassenegger & Pélissier, 1998). Moreover, mis-expression of *ZmMlp1* is not particularly biologically relevant at this stage. The temporally and spatially controlled expression pattern as discussed in chapter 3 could be indicative that the functional capabilities of *Zm*MLP1 are restricted, or that *Zm*MLP1 could have serious biochemical consequences if expressed in the wrong cell type or timeframe.

A more informative and reliable approach would be to analyse *ZmMlp1* from a loss-of function perspective. In order to achieve this, RNA interference (RNAi) technology was used to post-transcriptionally down-regulate *ZmMlp1* in a tissue specific manner. RNAi makes use of plant viral defence systems to target genes for silencing by degrading the mRNA in a sequence specific fashion, allowing targeted expression knock-down of

genes of interest (Carthew, 2001).

In order to characterise any observed phenotypes in the RNAi lines, a variety of techniques were used. Given the hypothesised function in seed development and nutrient translocation, seeds from the transgenic lines were measured with respect to overall seed size, weight and proportion of the seed occupied by starchy endosperm. This was achieved mainly by weighing individual seeds and image analysis with starch staining in order to differentiate tissue types by eye. Embryo and ESR development however, was assessed by fixation of developing seeds and histological analysis of seed sections in order to gain a cellular view of the effects of the RNAi construct.

Extensive molecular characterisation was also carried out on the transgenic lines to further investigate the transcriptional consequences of *ZmMlp1* knock-down. This was carried out by both qRT-PCR for a small number of genes of interest, together with a transcriptome wide approach through shotgun sequencing by RNAseq. While there are other approaches available for this kind of analysis such as tilling microarrays and EST sequencing, RNAseq offers a higher resolution, lower cost and more sensitive technology (Wang *et al.* , 2009; Nagalakshmi *et al.* , 2008; Wilhelm *et al.* , 2008), with qRT-PCR available for result confirmation.

Given the use of the synthetic *Zm*MLP1-GFP fusion construct in the previous chapter, genetic complementation experiments are also reported in this chapter which use reciprocal crosses to determine the functional status of the GFP transgenic lines. Furthermore, the synthetic nature of the GFP construct causes an inherent insensitivity to the RNAi construct, allowing further genetic complementation experiments to probe the individual parental gene function contributions and investigate the importance of the expression timeframe.

Using these approaches, the work in this chapter investigates the functional characteristics of *Zm*MLP1 and its importance in the developmental processes of plant reproduc-

tion. The data collected here is then analysed and discussed with reference to current evidence, raising new questions about the way in which the structure of *Zm*MLP1 facilitates its functionality.

Results

4.2 Functional investigation of *Zm*MLP1 by RNA interference

A construct was designed which contained a 297 bp fragment of the *ZmMlp1* gene, followed by an inverted repeat of the same fragment, separated by a spacer sequence (Figure 7.4, see 2.1.9). When expressed, this construct will then form a double stranded RNA (dsRNA) hairpin recognised by the Dicer protein which in turn cleaves the dsRNA into 20-25 nucleotide short interfering RNAs (siRNA) (Saurabh *et al.* , 2014; Pare & Hobman, 2007). These siRNAs then recruit the RISC complex which binds to complementary mRNA, cleaving it and targeting the cleavage products for degradation (Saurabh *et al.* , 2014; Redfern *et al.* , 2013; Wilson & Doudna, 2013; Riley *et al.* , 2012; Ender & Meister, 2010).

This RNAi construct was placed under the control of the native *ZmMlp1* promoter and used to generate the plasmid pBIOS 2755 (Figure 7.9). pBIOS 2755 was then transformed into A188 maize by Biogemma to generate 60 independent transgenic plant lines of the construct line T02297, which were selected for positive transformants containing only one transgene insertion (see 2.1.9). Seed stocks were then propagated by Biogemma through back-crossing with wild type pollen and the resulting seeds from twelve lines were transferred to Warwick for further analysis. Seeds were genotyped visually by use of the GFP marker gene (see 2.2.9) and both negative (WT) and positive transgenic T02297 seeds were then grown to maturity (see 2.1.1) and crossed by sibling pollen (see 2.1.5). Developing seeds at 5 DAP were then sampled from both positive and negative

plants, RNA extracted (see 2.2.2), cDNA synthesised (see 2.2.3) and the expression of *ZmMlp1* analysed by qRT-PCR (see 2.2.6). Expression analysis was able to identify five T02297+ sub-lines with significantly reduced *ZmMlp1* expression relative to the RNAi negative control plants (T02297-), which were then selected for further analysis (Figure 4.1). Considering that the T02297+ seeds were segregating at this point, the efficacy of the RNAi construct has almost certainly been under-represented in Figure 4.1, due to the presence of wild type mRNA in the samples.

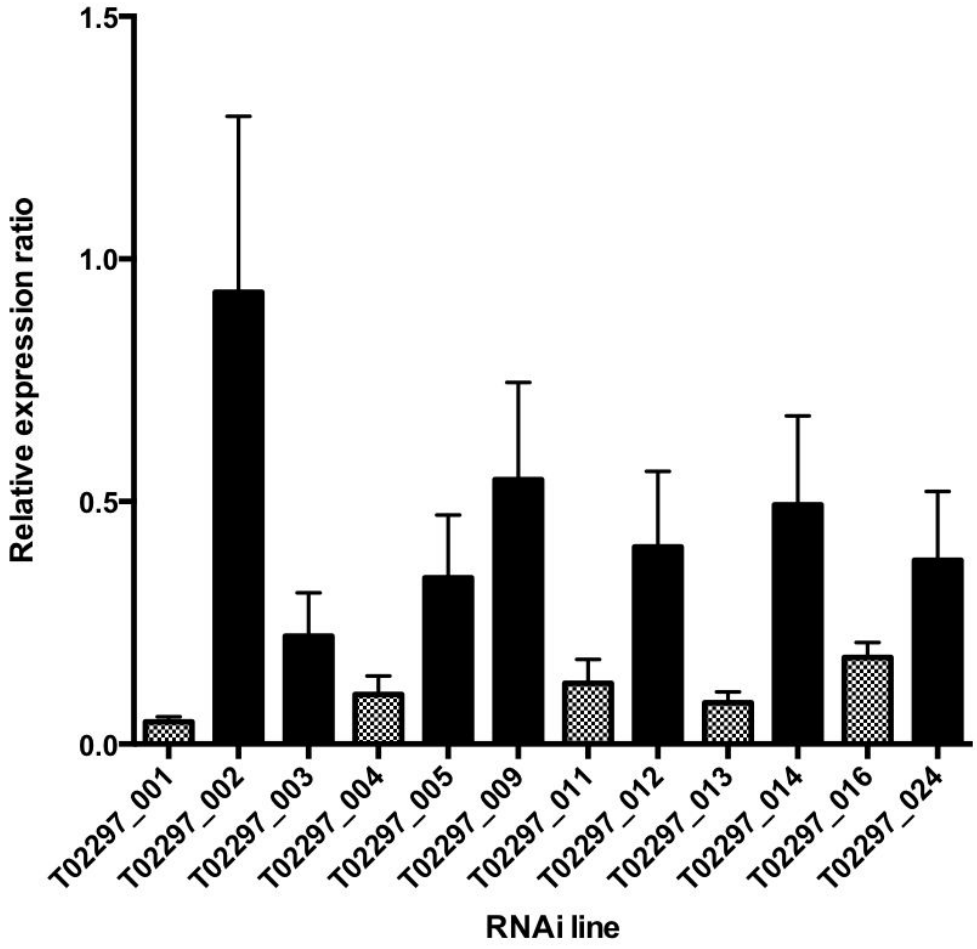

Figure 4.1: RNAi based down-regulation of ZmMlp1 expression. qRT-PCR analysis of 12 independent pro*Mlp1*:*Mlp1-RNAi* lines, analysing the expression level of *Mlp1* in whole developing seeds, sampled at 5 DAP. All data is relative to the ubiquitous gene *Gapdh* and then normalised to the relevant control lines. Bars in grey indicate a reduction in expression to less than 25% of the WT control. Error bars = propagated standard error of technical replicates.

4.3 Phenotypic characterisation of *ZmMlp1* RNAi lines

4.3.1 *Zm*MLP1 is required for normal mature seed morphology

In order to investigate the effect of reducing the expression level of *ZmMlp1*, plants from *ZmMlp1* RNAi (T02297+) sub-lines of interest (Figure 4.1), together with the corresponding wild type control lines (T02297-) were grown to maturity (see 2.1.1) and sib crossed (see 2.1.5). The seeds were then allowed to develop to maturity and dried for phenotypic analysis. Ears from wild type T02297- plants presented generally uniform seed size and shape across the central portion of the ear (Figure 4.2 (a, c and e)). However, ears from RNAi (T02297+) plants exhibited seeds that are squashed and shrivelled in appearance, indicating some level of seed collapse, sub-normal grain filling or developmental retardation (Figure 4.2 (b, d and f) (black arrows)).

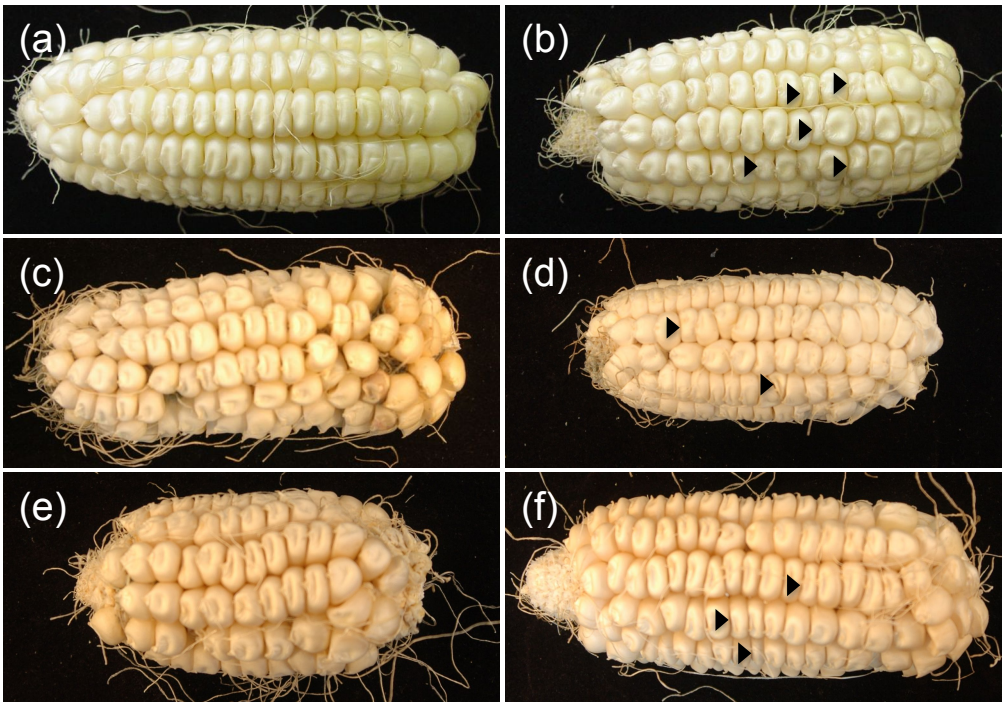

Figure 4.2: Dry seed analysis of pro*Mlp1*-RNAi (T02297) lines. Images show mature, dry ears photographed at the same magnification. (a, c and e) Line specific controls generated from the same transgenic event but containing no transgene (a) T02297_016^{-}, (c) T02297_001^{-} and (e) T02297_011^{-}. (b, d and f) Ears from plants that were genotyped for the presence of the RNAi construct (b) T02297_016^{+}, (d) T02297_001^{+}, (f) T02297_011^{+}. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants. Difference in shade of seeds due to photography effects.

Seeds were removed from the ear and examined individually to further characterise the phenotype. Isolated *ZmMlp1* RNAi seeds from 3 independent sub-lines demonstrate a striking morphological deformity compared to the wild type control (Figure 4.3 (top)). Wild type seeds show uniformity in size and morphology while *ZmMlp1* RNAi seeds exhibit a severely shrivelled and collapsed appearance and significantly reduced seed size. In order to confirm the visual phenotype, seed area was quantified for T02297_016, T02297_011 and T02297_001 by image analysis across 3 biological replicates for each sub-line (see 2.7.1). A statistically significant reduction in *ZmMlp1* RNAi (T02297+) seed area was observed for all sub-lines compared to the wild-type (T02297-) control

(Figure 4.3 (middle row) and Table 4.2). Additionally, grain weight was measured for T02297_016, T02297_011 and T02297_001 across 3 biological replicates for each subline (see 2.7.2), also yielding a statistically significant reduction in all three lines (Figure 4.3 (bottom) and Table 4.1).

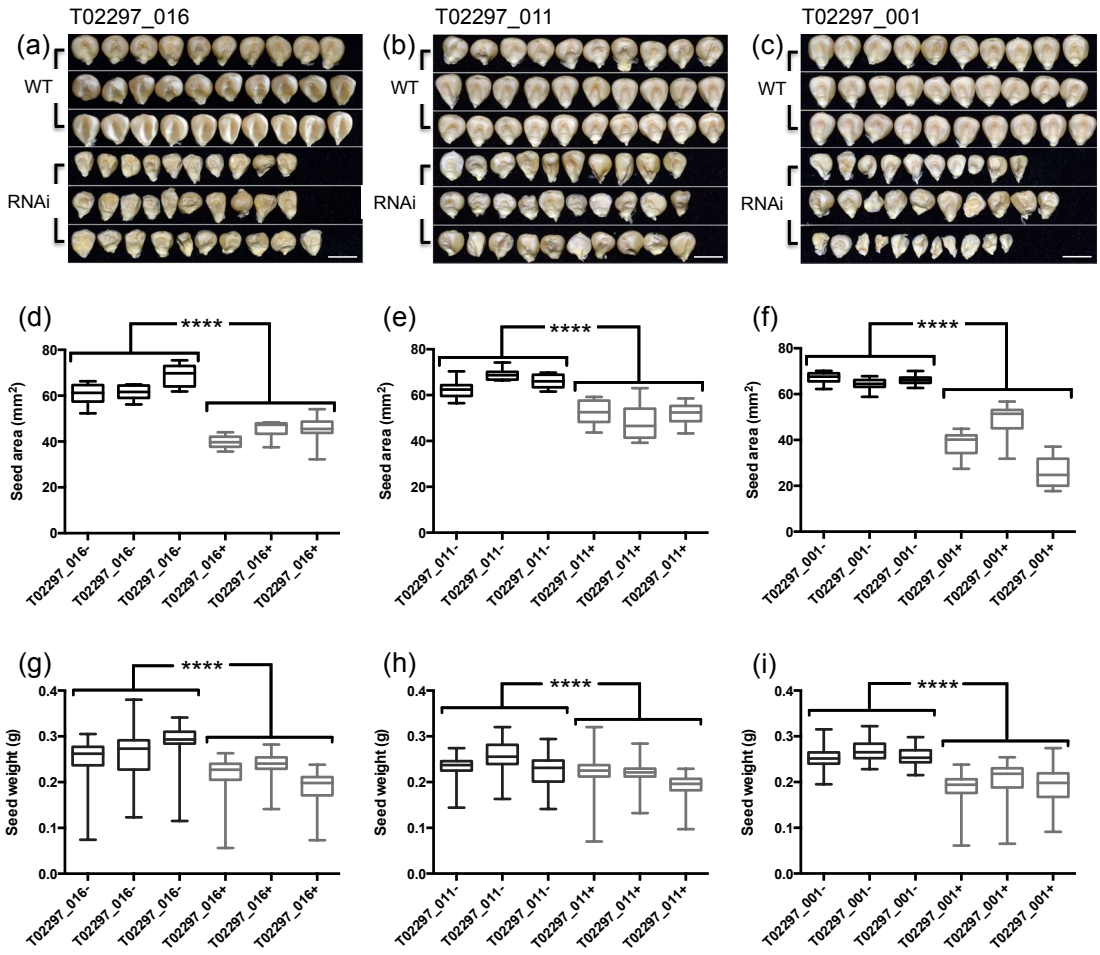

Figure 4.3: Dry seed analysis of pro*Mlp1*-RNAi lines. (a, d and g) T02297_016 (b, e and h) T02297_011 (c, f and i) T02297_001. (a-c) Images show mature, dry seeds photographed at the same magnification. Scale bars = 10 mm. (d-f) Box and whisker plot of dry seed area from pro*Mlp1*-RNAi lines. (g-i) Box and whisker plot of dry seed weight from pro*Mlp1*-RNAi lines. Black plots represent the WT (RNAi negative) control, grey plots represent the RNAi positive plants. All analyses taken in triplicate across 3 biological replicates for each sub-line. Seed data analysed by one-way ANOVA with multiple comparisons (Table 4.1 and Table 4.2) P value of < 0.0001 represented by ****. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants.

To investigate the underlying reasons behind the observed reduction in grain size and weight, internal seed morphology was analysed. T02297 seeds were bisected and the starch stained with iodine solution, allowing visual analysis and the quantification of

Table 4.1: One-way ANOVA multiple comparison test (Sidak) of mean dry seed weight in T02297 (MLP1 RNAi) lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene.

Line	Mean seed weight (g)	n	Mean Difference	95 % CI of MD	Significance
T02297_016-	0.265	290	0.0501	0.04392 to 0.05627	****
T02297_016+	0.214	332			
T02297_011-	0.236	369	0.0240	0.01883 to 0.02909	****
T02297_011+	0.212	572			
T02297_001-	0.257	373	0.0615	0.05585 to 0.06718	****
T02297_001+	0.195	363			

Table 4.2: One-way ANOVA multiple comparison test (Sidak) of mean dry seed area in T02297 (MLP1 RNAi) lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene.

MLP1 RNAi Line	Mean seed area (mm^2)	n	Mean Difference	95 % CI of MD	Significance
T02297_016-	63.67	30	20.07	16.11 to 24.04	****
T02297_016+	43.59	30			
T02297_011-	65.79	30	14.97	11.01 to 18.93	****
T02297_011+	50.82	30			
T02297_001-	65.91	30	28.29	24.33 to 32.25	****
T02297_001+	37.62	30			

seed endosperm proportion and seed cross-sectional area (see 2.7.3). The internal seed morphology was generally uniform in the wild type (T02297-) sub-lines, while the *ZmMlp1* RNAi (T02297+) sub-lines showed marked abnormalities (Figure 4.4 (top row)). The *ZmMlp1* RNAi seeds exhibited a highly variable morphology, with some seeds showing obvious defects in grain filling as well as scutellum which look abnormal. There was a reduction in the percentage of seed occupied by starchy endosperm in all three sub lines, though the difference was statistically significant only in T02297_016 (Figure 4.4 (middle row) and Table 4.3). However, similar to the observed reduction in seed area, the reduction in cross-sectional area of RNAi seeds was found to be statistically significant when compared to the wild type seeds (Figure 4.4 (bottom row) and Table 4.4).

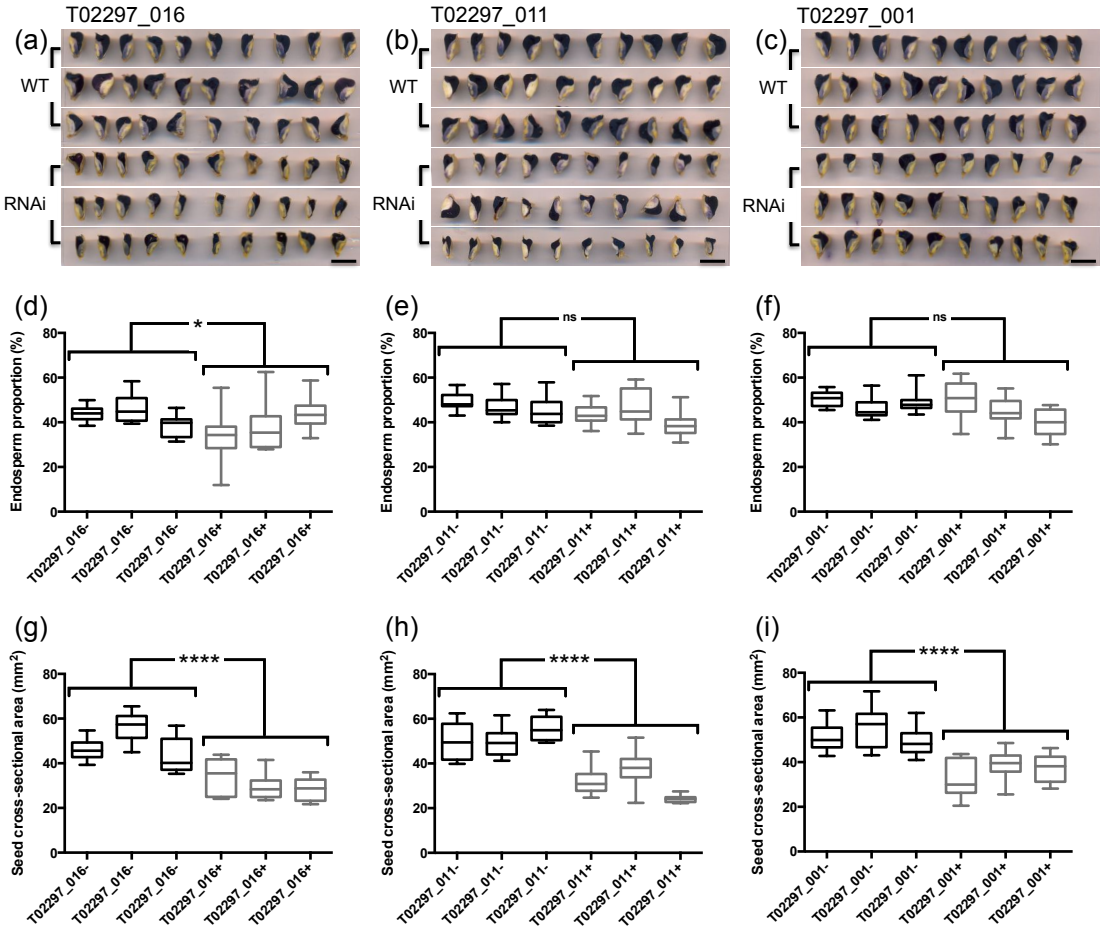

Figure 4.4: Dry seed analysis of pro*Mlp1*-RNAi lines. (a, d and g) T02297_016 (b, e and h) T02297_011 (c, f and i) T02297_001. (a-c) Images show mature, dry seeds from independent T02297- plants (top 3 rows) and T02297+ plants (bottom 3 rows) which have been bisected across the central vertical axis and stained for starch with iodine solution (black precipitate). Scale bars = 10 mm. (d-f) Box and whisker plot of proportion of seed cross-sectional area composed of starchy endosperm in T02297 (pro*Mlp1*-RNAi) lines. (g-i) Box and whisker plot of seed cross-sectional area from T02297 (pro*Mlp1*-RNAi) lines. Black plots represent the WT (RNAi negative) control, grey plots represent the RNAi positive plants. All measurements taken in triplicate across 3 biological replicates. Seed data analysed by one-way ANOVA with multiple comparisons (Table 4.3 and Table 4.4) P values of <0.05 and < 0.0001 represented by * and **** respectively. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants.

Table 4.3: One-way ANOVA multiple comparison test (Sidak) of seed endosperm proportion in RNAi lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene.

MLP1 RNAi Line	Mean endosperm proportion (%)	n	Mean Difference	95 % CI of MD	Significance
T02297_016-	42.87	30	4.482	0.02092 to 8.943	*
T02297_016+	38.39	30			
T02297_011-	47.09	30	4.044	-0.4172 to 8.505	ns
T02297_011+	43.05	30			
T02297_001-	48.50	30	3.403	-1.058 to 7.864	ns
T02297_001+	45.09	30			

Table 4.4: One-way ANOVA multiple comparison test (Sidak) of seed cross-sectional area in RNAi lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene.

MLP1 RNAi Line	Mean cross-sectional area (mm^2)	n	Mean Difference	95 % CI of MD	Significance
T02297_016-	48.71	30	18.18	13.47 to 22.89	****
T02297_016+	30.53	30			
T02297_011-	51.83	30	20.51	15.80 to 25.21	****
T02297_011+	31.32	30			
T02297_001-	52.09	30	16.15	11.45 to 20.86	****
T02297_001+	35.94	30			

4.3.2 *Zm*MLP1 is involved in early embryo pattern formation

The expression pattern as described in chapter 3 indicated that *Zm*MLP1 functions during the developmental stages of the endosperm and early embryo formation. Since the homologue *Zm*MEG1 functions in a developmental role, it was hypothesised that *Zm*MLP1 may operate in a similar capacity. In order to investigate any developmental consequences of reducing *ZmMlp1* expression, *ZmMlp1* RNAi (T02297+) plants were grown to maturity (see 2.1.1), sib-crossed (see 2.1.5) and sampled at 5 DAP for analysis by *in situ* hybridisation (see 2.5.1, 2.5.2 and 2.5.4).

Three different RNA probes were used to label the ESR (*AE1* (AJ011559.1) (Magnard *et al.* , 2000) and *ESR2* (NM_001112006.1) (Opsahl-Ferstad *et al.* , 1997)) and the protoderm of the proembryo (*Lipid Transfer Protein 1 (LTP1)* (GRMZM2G101958) in MLP1 RNAi (T02297+) and wild-type (T02297-) seed sections. Sense strand control probes were also tested for each probe, all showing no signal localisation (not shown). At this developmental time-point, no differences were observed in the ESR morphology or probe hybridisation between *ZmMlp1* RNAi seeds and the wild type, however abnormalities were observed in the dimensional morphology and patterning of the proembryo (Figure 4.5). The suspensors of the embryos in the *ZmMlp1* RNAi seeds appear to be shorter, wider and contain more cells than the wild type. Additionally, the early embryo proper appears to be reduced in size and in some cases morphologically abnormal. The accumulation of the *LTP1* probe was also observed to cover a smaller region of the protoderm in the *ZmMlp1* RNAi seeds (Figure 4.5 (bottom)), which aligns well with the previous observation of *Zm*MLP1 localising to the embryo protoderm (discussed in chapter 3).

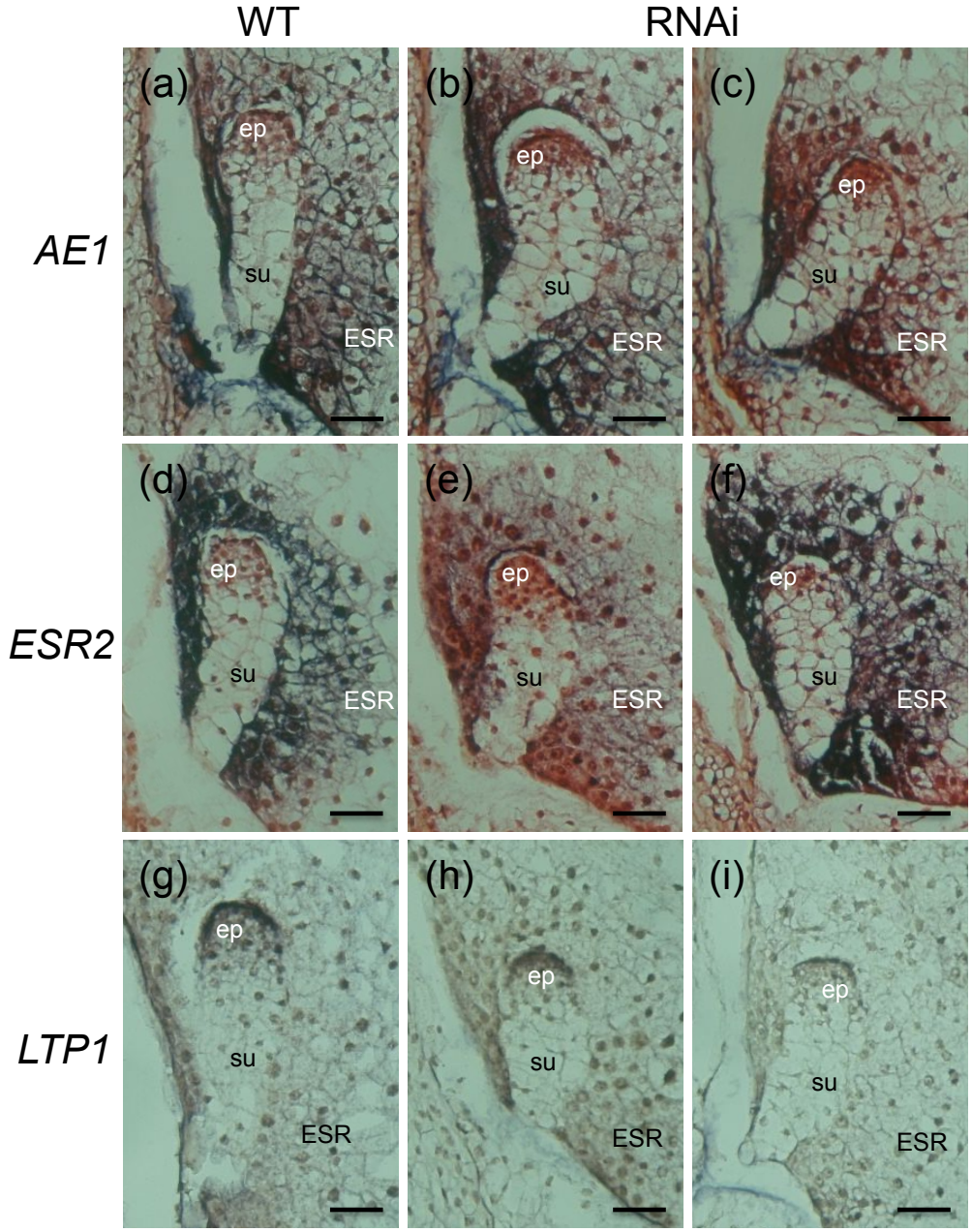

Figure 4.5: *In situ* hybridisation analysis 6 dDAP seeds from wild type and transgenic (T02297) plants. (Top row) RNA probe for *Androgenic Embryo 1* (*AE1*), (middle row) probe for *Embryo Surrounding Region 2* (*ESR2*), (Bottom row) probe for *Lipid Transfer Protein 1* (*LTP1*). (Left panel) Wild type seed samples, (centre and right panel) MLP1 RNAi phenotype. Probe localisation indicated by dark purple precipitate, counterstaining with safranin (red-brown colour). ESR = Embryo surrounding region, ep = embryo proper, su = suspensor. All probes were tested against the sense control probe for spurious signal before use (not shown). Scale bars = 50 µm

In order to confirm and quantify the embryo phenotype observed in Figure 4.5, image analysis was carried out to measure the dimensional changes to the embryo in the *ZmMlp1* RNAi (T02297+) lines compared to the wild type (T02297-) (see 2.7.4). The total proembryo length was found to be statistically significantly reduced in the *ZmMlp1* RNAi lines compared to the wild type (Figure 4.6 (d) and Table 4.5). The total area occupied by suspensor cells was then measured and the difference found to be statistically insignificant between *ZmMlp1* RNAi and the wild type , though the suspensor length was significantly reduced leading to a significantly increased mean suspensor width as represented by the suspensor area to length ratio (Figure 4.6 (a, b and c) and Table 4.5). This short, fat suspensor (SFS) phenotype was accompanied by a significant reduction in the area of the embryo proper, which led to an increase in the proportion of the embryo occupied by the suspensor as indicated by the increased suspensor area to embryo proper area ratio (Figure 4.6 (e and f) and Table 4.5). Together the SFS phenotype and reduced embryo proper (REP) phenotype demonstrate a significant alteration to the normal embryo pattern formation and since this leads to a defective kernel as previously discussed, this phenotype shall be referred to as defective embryo patterning (DEP).

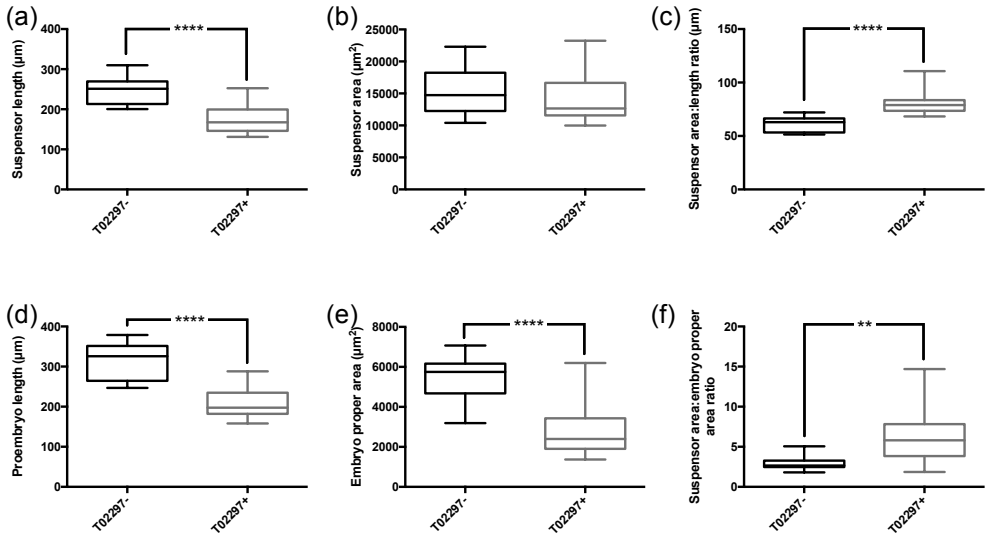

Figure 4.6: Dimensional analysis of proembryos in developing *ZmMlp1* RNAi (T02297) seeds at 6 DAP. (a and b) Box and whisker plots of mean suspensor length (µm) and area (µm^2) respectively. (c) Box and whisker plot of mean suspensor width represented as suspensor area:length ratio (µm). (d and e) Box and whisker plots of total proembryo length (µm) and embryo proper area (µm^2) respectively. (f) Proembryo suspensor proportion represented as suspensor area:embryo proper area ratio. Black plots represent the wild type (T02297-) control, grey plots represent the RNAi (T02297+) plants. Seed data analysed by two-tailed Student's t-test (Table 4.5) P values of <0.01 and < 0.0001 represented by ** and **** respectively. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants.

4.3.3 Molecular analysis of the DEP phenotype

Using phenotypic analysis, it has been established that *ZmMlp1* is an important gene for normal embryo and endosperm development. However, this does not reveal any information how *Zm*MLP1 fits into the molecular mechanisms that regulate the developmental processes in the seed. To investigate the genetic interactions of *ZmMlp1*, the *ZmMlp1* RNAi lines were subjected to global expression analysis by whole transcriptome shotgun sequencing (RNA-seq). RNA-seq is a relatively recent technology which makes use of advancements in high-throughput sequencing to rapidly and accurately

Table 4.5: Two-tailed Student's t-test of embryo dimensions in RNAi (T02297) lines. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants.

MLP1 RNAi Line	Mean suspensor length (µm)	n	Mean Difference	95 % CI of MD	Significance
T02297-	245.5	13	-68.72	-94.64 to -42.80	****
T02297+	176.8	18			
	Suspensor area (μm^2)				
T02297-	15175	13	-596.3	-3618 to 2426	ns
T02297+	14579	18			
	Suspensor area:length ratio (µm)				
T02297-	61.24	13	20.26	12.85 to 27.67	****
T02297+	81.50	18			
	Proembryo length (µm)				
T02297-	314.0	13	-102.1	-134.0 to -70.22	****
T02297+	211.9	18			
	Embryo proper area (μm^2)				
T02297-	5421.0	13	-2616.0	-3531 to -1701	****
T02297+	2805.0	18			
	Suspensor area:Embryo proper area				
T02297-	2.88	13	3.41	1.410 to 5.400	**
T02297+	6.28	18			

sequence libraries of cDNA fragments generated from an RNA population. This allows identification and quantification of expressed transcripts down to a single nucleotide resolution with advantages in sensitivity and cost over traditional technologies such as tiling microarray and EST sequencing (Wang *et al.* , 2009; Nagalakshmi *et al.* , 2008; Wilhelm *et al.* , 2008).

In order to control for natural and environmental variation in gene expression, it was important to select *ZmMlp1* RNAi and wild type samples at the same developmental stage. In order to achieve this, plants from T02297- and T02297+ sub-lines were grown to maturity (see 2.1.1), crossed by sibling pollen (see 2.1.5) and endosperms isolated at 5 DAP by micro-dissection. RNA was then extracted (see 2.2.2) and used to generate endosperm enriched cDNA libraries (see 2.2.11). As a further control, cDNA libraries were initially analysed by qRT-PCR to select samples with similar expression levels of the gene *ZmAL9*, an aleurone specific gene which is a useful marker of endosperm developmental stage (Gómez *et al.* , 2009). Selected cDNA libraries were then sequenced by Illumina sequencing.

The sequencing results were then analysed by Biogemma to determine the genes that were significantly altered in expression (at least 1-fold change, p value < 0.05) in the RNAi endosperms and identify the gene identifier codes from the Maize Genetics and Genomics Database (MaizeGDB). These gene IDs were then used to interrogate the MaizeGDB to determine if the genes had been characterised, or if there were characterised homologous genes in the genomes of *Arabidopsis* and Rice, using the Phytozome annotations. This data was then used to generate putative categories for the genes in maize which were then clustered into the broad groups of: CRPs, small peptides, metabolism, signalling, transcription, and transport and storage. Interestingly, the down-regulated genes were strongly enriched for CRPs and genes involved in metabolic processes (Figure 4.7), reflecting the highly metabolic nature of the ESR. The

down-regulation of so many CRPs could potentially be indicative of peptide based signalling processes down-stream from *Zm*MLP1, or potentially CRPs that interact with *Zm*MLP1 in a complex. This could be due to regulatory negative-feedback loops which keep the signalling process in balance, a further indication of which is the alteration in expression of transcription factors (Figure 4.7).

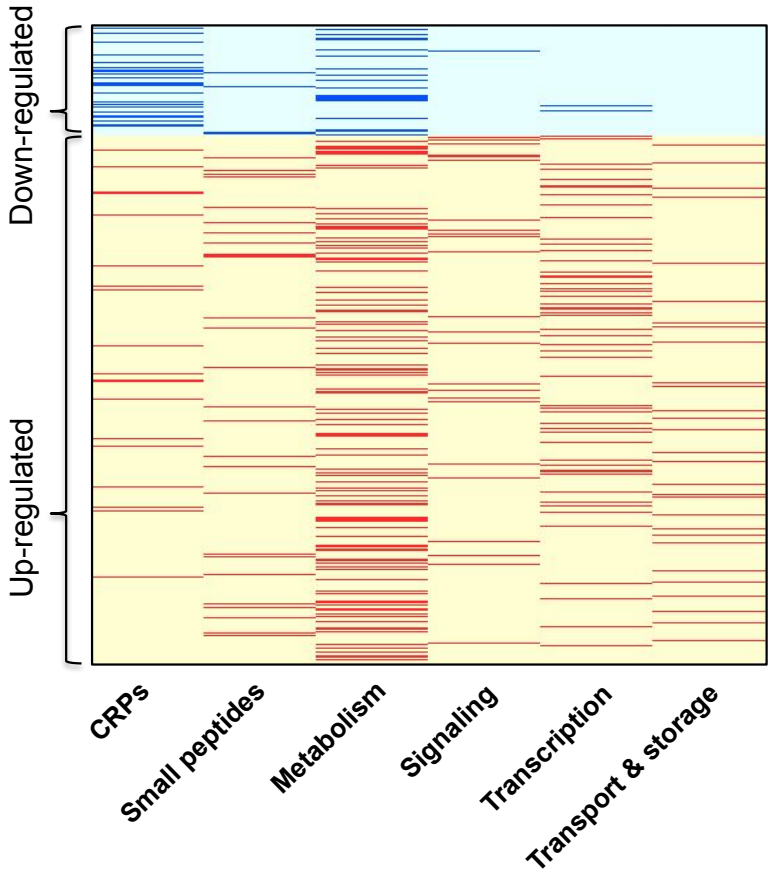

Figure 4.7: Differential gene expression analysis of *ZmMlp1*-RNAi plants by RNA-seq. A differential expression heat map showing the most significantly up- and down-regulated genes in the endosperm of *ZmMlp1* RNAi (T02297+) endosperms relative to the wild-type control (T02297-). Genes showing more than 1-fold change in expression were clustered into broad functional categories and colour coded according to up-regulation (red) or down-regulation (blue).

The up-regulated genes were also enriched in metabolic genes, as expected in a metabol-

ically active tissue. In contrast however, there was significant enrichment in the other categories of small peptides, signalling components, transcription factors, and genes associated with transport and storage (Figure 4.7). These data suggest that *Zm*MLP1 is involved in a signalling pathway that could involve up- or down-stream transcriptional regulation. Some of the transport and storage genes may also be either directly or indirectly involved in this hypothetical signalling pathway, potentially accounting for the observed reduction in seed biomass in *ZmMlp1*-RNAi lines.

Particular genes of interest with decreased or increased expression can be found in Tables 4.6 and 4.7 respectively. As expected, the most highly down-regulated gene in the *ZmMlp1*-RNAi lines was *ZmMlp1*, showing consistent down-regulation across the replicates. Interestingly, two genes from the *ZmBAP2* family (GRMZM2G008403 and GRMZM2G027472) were down-regulated, a family thought to be involved in pathogen defence in the BETL (Serna *et al.* , 2001). Two other down-regulated genes, GRMZM2G087766 and GRMZM2G046086 were shown to encode the characterised genes *ZmEBE-2* and *ZmESR1* respectively. *ZmESR1* is an ESR specific gene thought to be involved in pathogen defence (Bonello *et al.* , 2000), while *ZmEBE-2* interestingly shares similar expression characteristics as *ZmMlp1*, being expressed in the central cell, and is thought to have a role in the early developmental stages of the endosperm (Magnard *et al.* , 2003). Several unidentified CRPs were also found to be down-regulated (see Table 4.6), most of which show exclusive expression in the ovule and early seed in published microarray data (Winter *et al.* , 2007; Waters *et al.* , 2011; Davidson *et al.* , 2011).

Table 4.6: Putative functions for selected down-regulated genes as a consequence of *ZmMLP1* RNAi.

Gene ID	**Name**	**Expression**	**Putative function**	**Fold change**	**References**
GRMZM2G145466	*ZmMLP1*	ESR	Embryo patterning	-2.79	
GRMZM2G489599	Novel CRP	Unknown	Unknown	-1.65	
GRMZM2G008403	*ZmBAP* family	BETL	Pathogen defense	-1.39	(Serna *et al.* , 2001)
GRMZM2G306105	Novel CRP	Leaves	Unknown	-1.35	(Davidson *et al.* , 2011; Winter *et al.* , 2007)
GRMZM2G483275	Novel CRP	Ovule & Early seed	Unknown	-1.30	(Davidson *et al.* , 2011)
GRMZM2G074293	Novel CRP	Ovule & Early seed	Unknown	-1.29	(Davidson *et al.* , 2011)
GRMZM2G001827	Novel CRP	Unknown	Unknown	-1.29	
GRMZM2G483273	Novel CRP	Ovule & Early seed	Unknown	-1.26	(Davidson *et al.* , 2011; Winter *et al.* , 2007)
GRMZM2G489627	Novel CRP	Ovule & Early seed	Unknown	-1.22	(Davidson *et al.* , 2011)
GRMZM5G815477	Novel CRP	Early seed	Unknown	-1.20	(Davidson *et al.* , 2011; Winter *et al.* , 2007)
GRMZM2G087766	*ZmEBE2*	CC and BETL	Development?	-1.18	(Magnard *et al.* , 2003)
GRMZM2G046086	*ZmESR1*	ESR	Pathogen defense?	-1.03	(Bonello *et al.* , 2000)
GRMZM2G027472	*ZmBAP* family	BETL	Pathogen defense	-1.03	(Serna *et al.* , 2001)

Table 4.7: Putative functions for selected up-regulated genes as a consequence of *ZmMLP1* RNAi.

Gene ID	Name	Expression	Putative function	Fold change	References
GRMZM2G087094	*At*NDR1/HIN1-like1 homologue	Early seed	Virus defence	2.43	(Davidson *et al.* , 2011)
GRMZM2G039538	*AtALE1* homologue	Early seed	Embryo development (*At*)	2.10	(Davidson *et al.* , 2011; Winter *et al.* , 2007)
GRMZM2G161641	*AtAAP* homologue	Ovule & Early seed	Amino acid transporter	1.89	(Davidson *et al.* , 2011)
GRMZM2G131177	*At*Subtilase homologue	Ovule, early seed & silks	Embryo development (*At*)	1.81	(Davidson *et al.* , 2011)
GRMZM2G089713	*ZmSH-1*	Endosperm	Sucrose synthesis/degradation	1.80	(Davidson *et al.* , 2011; Winter *et al.* , 2007)
GRMZM2G095725	*AtCWIN2* homologue	Early seed	Sucrose inversion	1.74	(Davidson *et al.* , 2011)
GRMZM2G057159	*AtALE1* homologue	Ovule & Early seed	Embryo development (*At*)	1.62	(Davidson *et al.* , 2011)
GRMZM2G368827	*Os*MtN3 family homologue	Early seed	Sugar transporter	1.56	(Davidson *et al.* , 2011; Winter *et al.* , 2007)
GRMZM2G158240	*At*PME61 homologue	Ovule & Early seed	Cell wall extension	1.51	(Davidson *et al.* , 2011)
GRMZM5G872141	*Os*MtN3 family homologue	Endosperm	Sugar transporter	1.41	(Davidson *et al.* , 2011)
GRMZM2G478876	*At*Protein kinase superfamily homologue	Embryo, ovule & early seed	Signalling?	1.31	(Davidson *et al.* , 2011; Waters *et al.* , 2011)
GRMZM2G043584	*OsCLAVATA1* precursor homologue	Embryo, ovule & early seed	Shoot apical meristem formation	1.30	(Davidson *et al.* , 2011; Waters *et al.* , 2011)
GRMZM2G126765	*At*Phosphotyrosine phosphatase homologue	Ovule & Early seed	Signalling?	1.23	(Davidson *et al.* , 2011)
GRMZM2G363552	*AtALE1* homologue	Early seed	Embryo development (*At*)	1.23	(Davidson *et al.* , 2011; Winter *et al.* , 2007)
GRMZM2G017164	*AtRLK-1* homologue	Ovule, early seed & silks	Signalling?	1.22	(Davidson *et al.* , 2011)
GRMZM2G136106	*AtPME61* homologue	Embryo, ovule & early seed	Cell wall extension	1.22	(Davidson *et al.* , 2011; Waters *et al.* , 2011)
GRMZM2G115658	*AtZIFL1* homologue	Embryo, ovule & early seed	Ion transporter?	1.19	(Davidson *et al.* , 2011)
GRMZM2G169681	*OsCLAVATA1* precursor homologue	Ovule & Early seed	Shoot apical meristem formation	1.12	(Davidson *et al.* , 2011)
GRMZM2G055684	*AtALE1* homologue	Ovule & Early seed	Embryo development (*At*)	1.08	(Davidson *et al.* , 2011)

The up-regulated gene compendium also presented some interesting candidates, many of which are expressed exclusively in the ovule and the early seed according to the microarray data (Table 4.7). The *ZmSH-1* gene (GRMZM2G089713) was found to be significantly up-regulated and encodes the sucrose synthase enzyme, which despite its name, is primarily involved in sucrose degradation in the process of starch synthesis (Cobb & Hannah, 1988). *ZmSH-1* is the deficient enzyme in the *Shrunken-1* mutant, which presents a similar phenotype to the *ZmMlp1*-RNAi lines. This could indicate that *Zm*MLP1 has a pathway connection to the starch synthesis mechanisms in the endosperm, or could be caused by off-target effects of the RNAi construct. Two other genes with homology to rice sugar transporters (GRMZM2G368827 and GRMZM5G872141) were also found to be up-regulated, adding to the evidence that *Zm*MLP1 has a knock-on effect in sugar partitioning in the seed. Furthermore, the up-regulated gene GRMZM2G095725, encodes a homologue for the *Arabidopsis* cell wall invertase, *At*CWIN2, expressed in the early seed of maize. This could be an unidentified maize invertase, suggesting a glucose signalling role in the *Zm*MLP1 pathway, although a rise in glucose levels would usually be associated with an increase in growth. However, there remains the possibility that an associated invertase inhibitor which has not yet been identified could also be up-regulated in the ESR, shielding the developing embryo from glucose mediated signalling, similar to the actions of *Zm*INVINH (Bate *et al.* , 2004).

The most interesting differentially expressed genes however, are four up-regulated subtilisin-like genes with homology to the *Arabidopsis abnormal leaf shape1 (AtALE-1)* (see Table 4.7). These genes are expressed in the ovule and early seed of maize, similar to the expression of *AtALE-1* which localises to the region of the endosperm surrounding the early embryo (Tanaka *et al.* , 2001). *AtALE-1* encodes a serine protease and has been shown to be involved in embryo development, required in *Arabidopsis* for formation of

the embryonic epidermis (Tanaka *et al.* , 2001).

Additionally, two homologues for the rice *OsCLAVATA1* gene were also up-regulated in response to *ZmMlp1* knock-down (see Table 4.7). CLAVATA1 (CLV1) is a receptor protein kinase which regulates the mitogen-activated protein kinase (MAPK) pathway, involved in the formation of the shoot and floral apical meristems in *Arabidopsis (Clark* et al. *, 1995; Betsuyaku* et al. *, 2011).* Significantly though, the subtilisin-like genes up-regulated in *ZmMlp1*-RNAi lines bear significant homology to another *Arabidopsis* subtilase, *At*SBT5.4, over-expression of which has been shown to cause a clavata-like phenotype through interaction with the WUSCHEL transcription factors (Liu *et al.* , 2009). The clavata-like phenotypes observed have wide ranging effects on the plant with alterations in development of roots, shoot apex meristems, petal development and more interestingly, the size and morphology of the siliques. This alignment in effect on reproductive development presents the possibility that an unidentified clavata-like pathway may be operating in the developmental processes of the early maize embryo, potentially regulated by a subtilase family protein, any component of which could be directly or indirectly interacting with *Zm*MLP1.

To further investigate the molecular consequences of reducing the expression of *ZmMlp1*, expression analysis was carried out on genes specific for different compartments within the endosperm. Endosperm specific cDNA libraries from wild type (T02297-) and *ZmMlp1* RNAi (T02297+) seeds were subjected to qRT-PCR (see 2.2.6), analysing the expression of the aleurone specific gene: *ZmAl9*, the ESR specific genes: *ZmESR1*, *ZmESR6* and *ZmMlp1* and the BETL specific genes: *ZmMeg1* and *ZmMrp1* (Figure 4.8 and Table 4.8). Interestingly, endosperms at similar developmental stages as demonstrated by *ZmAl9* expression levels showed no significant change in expression of ESR specific genes between *ZmMlp1* RNAi and the wild type (with the expected exception of *ZmMlp1*) (Figure 4.8 (a-d) and Table 4.8). These data combined with no

observed ESR phenotype suggest that *Zm*MLP1 does not have a developmental role in the ESR and therefore *Zm*MLP1 is not functionally homologous to *Zm*MEG1. As suggested by the expression characterisation in chapter 3 and the DEP phenotype, the lack of effect on the ESR by the knock-down of *ZmMlp1*, indicates that the functional role of *Zm*MLP1 is in cells exterior to the ESR, most likely in the developing embryo. However, the expression analysis of BETL specific genes showed a significant reduction in the expression of both *ZmMeg1* and *ZmMrp1* (Figure 4.8 (e and f) and Table 4.8), which could indicate that *Zm*MLP1 has a role in the BETL.

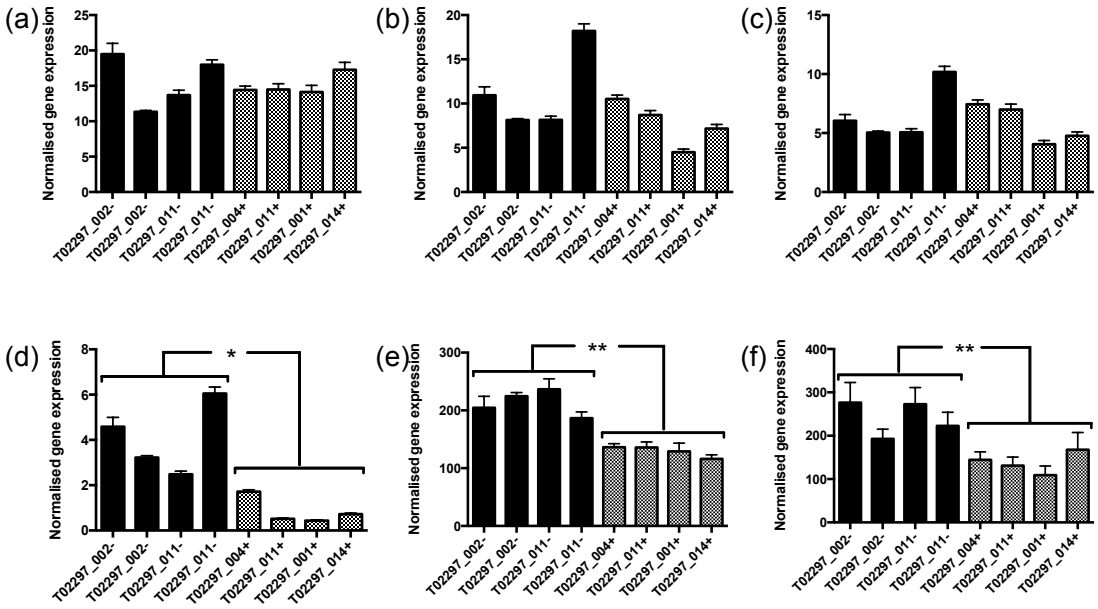

Figure 4.8: Expression analysis of endosperm compartment specific genes by qRT-PCR. WT (black) and RNAi (grey) samples were analysed in quadruplicate at 5 DAP from isolated endosperms, across 6 different target genes - (a) *ZmAl9*, (b) *ZmEsr1*, (c) *ZmEsr6*, (d) *ZmMlp1*, (e) *ZmMeg1* and (f) *ZmMrp1*. Expression data was analysed by unpaired, two tailed Student's t-test with Welches correction (Table 4.8). P values of <0.05 and <0.01 by * and ** respectively. P values <0.05 are not indicated. Error bars indicate standard error of technical replicates. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants.

Table 4.8: Students t-test (Welches) of gene expression levels for compartment specific genes in developing RNAi seeds (Line T02297). (+) Transgenic sibling plants; (-) Non-transgenic sibling plants as determined by PCR for BASTA resistance marker gene.

Gene	**Genotype**	**Mean Expression Level**	**n**	**Mean Difference**	**95 % CI of MD**	**Significance**
ZmAl9	T02297-	15.63	4	0.5448	-6.227 to 5.137	ns
	T02297+	15.09	4			
ZmEsr1	T02297-	11.36	4	3.268	-10.75 to 3.494	ns
	T02297+	7.73	4			
ZmEsr6	T02297-	6.587	4	0.771	-4.507 to 2.964	ns
	T02297+	5.815	4			
ZmMlp1	T02297-	4.086	4	3.237	-5.613 to -0.8617	*
	T02297+	0.8487	4			
Zm-Meg1	T02297-	213.1	4	116.9	-116.9 to -50.63	**
	T02297+	129.4	4			
Zm-Mrp1	T02297-	241.3	4	103.1	-164.1 to -42.14	**
	T02297+	138.2	4			

4.4 Genetic complementation of *ZmMlp1* RNAi by *Zm*MLP1-GFP

In chapter 3, a *Zm*MLP1-GFP fusion protein was used to localise expression of *Zm*MLP1 in maize seeds. This construct was comprised of a GFP sequence fused to a synthetic *ZmMlp1* sequence, which differs significantly in nucleotide sequence from the endogenous *ZmMlp1* gene (Figure 7.13). Due to the sequence divergence, the synthetic *ZmMlp1* should be insensitive to the RNAi mediated silencing, which relies on a high degree of sequence similarity (Miller *et al.* , 2003; Brummelkamp *et al.* , 2002). However, due to the relative size of the GFP tag, the *Zm*MLP1-GFP fusion protein could be functionally disrupted. To ascertain whether *Zm*MLP1-GFP is functional, a genetic complementation experiment was performed to determine if the presence of the *Zm*MLP1-GFP protein was able to recover the *ZmMlp1* RNAi phenotype.

An interesting characteristic of the expression pattern of *ZmMlp1* was the expression of the protein in the central cell before fertilisation. This suggests that the protein has a functional role either before fertilisation or during the very early developmental stages immediately after fertilisation. This central cell expression also complicates the genetic complementation experiment, since the orientation of the cross will dictate at what stage the synthetic gene is delivered. If the synthetic *ZmMlp1* is delivered from the female gamete, it will be present in the central cell. However, if the synthetic gene arrives from the male gamete, there would be a delay in its expression which could cause failure of complementation if the function of *Zm*MLP1 is crucial either before or immediately after fertilisation.

In order to determine if the maternal expression of *ZmMlp1* is functionally critical, reciprocal crosses were carried out using *ZmMlp1* RNAi (T02297) and *ZmMlp1*-GFP (T02492) lines together with reciprocal wild type control crosses. Seeds from reciprocal

and control crosses were sampled at 15 DAP, stained for starch accumulation in the endosperm and the proportion of seed area containing starch was quantified (see 2.7.3). The seeds from *ZmMlp1* RNAi (T02297) sibling and reciprocal A188 back-crosses, were found to exhibit a reduced or delayed starch accumulation in the seed compared to the *ZmMlp1*-GFP (T02492) sibling and reciprocal A188 back-crosses (Figure 4.9 and Table 4.9). Interestingly, recovery of the *ZmMlp1* RNAi phenotype was observed when the synthetic *ZmMlp1* was delivered from the female gamete in the GFP x RNAi cross, while the reciprocal cross was unable to complement fully (Figure 4.9 and Table 4.9).

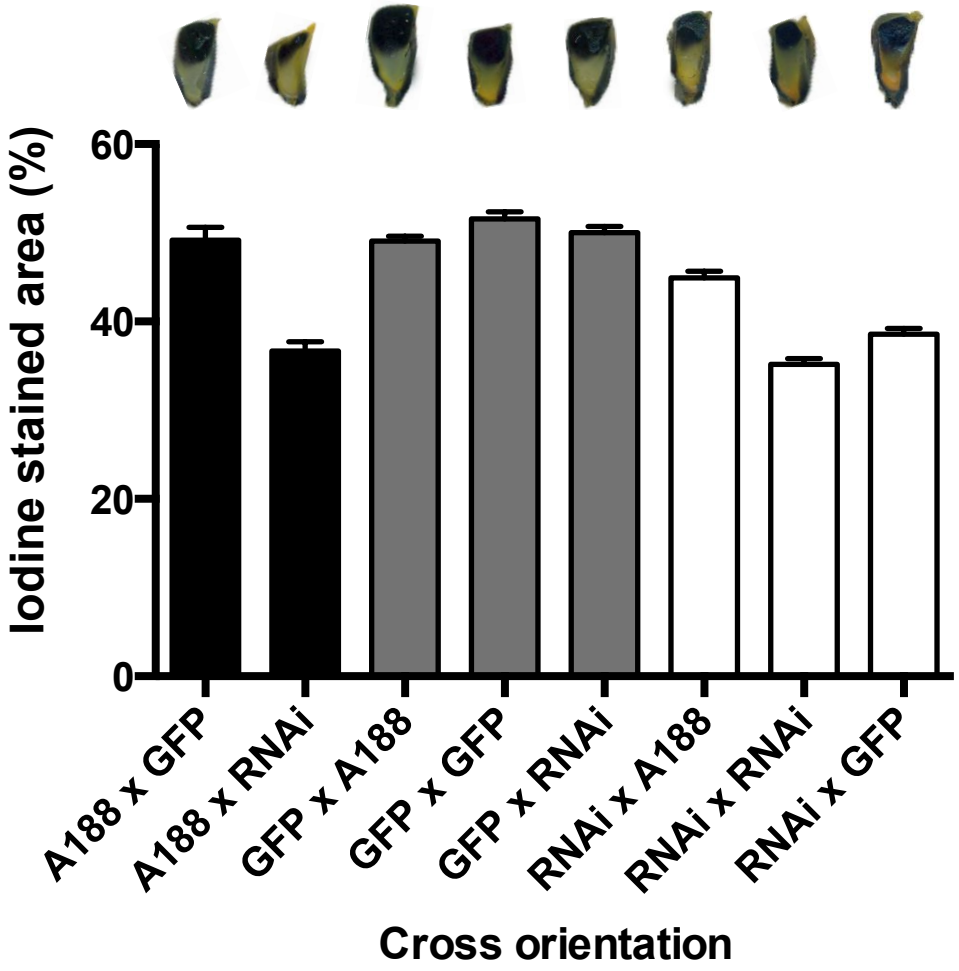

Figure 4.9: Genetic complementation test between pro*Mlp1*:RNAi and *Zm*MLP1-GFP. (Top) Example 15 DAP seed cross-sections stained with iodine generated from the crosses. (Bottom) Area of seed cross-section containing starch as indicated by iodine stain. Data pooled from independently crossed biological replicates and analysed by one-way ANOVA with multiple comparisons (Table 4.9) Error bars indicate standard error of the mean.

To further characterise the genetic complementation, seeds from reciprocal crosses between *ZmMlp1* RNAi and *Zm*MLP1-GFP were allowed to develop to maturity, bisected and analysed for seed cross-sectional area and starch proportion (see 2.7.3). As previously observed, the RNAi line (T02297) showed a significant decrease in seed

cross-sectional area (Figure 4.10 (left)), but no significant alteration to endosperm seed proportion (Figure 4.10 (right)). The *Zm*MLP1-GFP line was able to complement the *ZmMlp1* RNAi and recover the phenotype when delivered through the maternal line but not when delivered from the male (Figure 4.10 (left) and Table 4.10). Combined with the data from Figure 4.9, this also suggests that the reduction in starch proportion as seen at 15 DAP could be due to a delay in starch synthesis, rather than an absolute reduction in endosperm proportion, since this seems to recover by the mature stages of the seed.

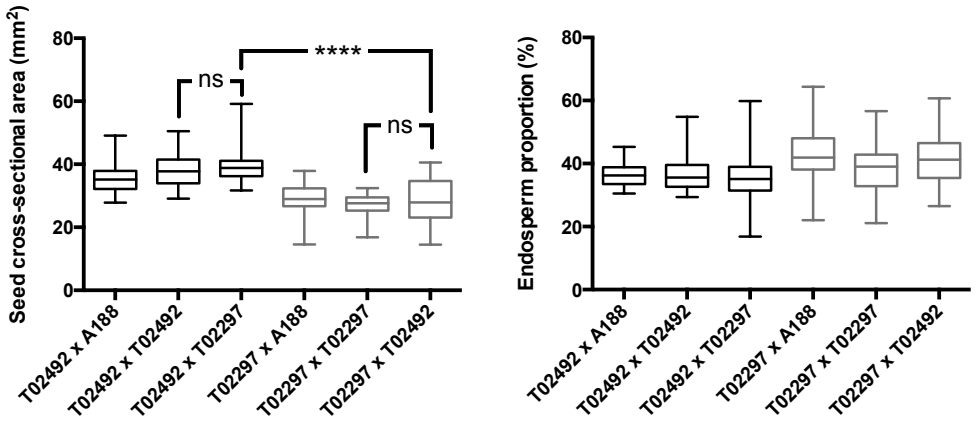

Figure 4.10: Genetic complementation analysis in reciprocal crosses of pro*Mlp1*:RNAi and *Zm*MLP1-GFP. (Left) Box and whisker plot of mature seed cross-sectional areas from *ZmMlp1* RNAi and *ZmMlp1*-GFP. (Right) Box and whisker plot of mature seed endosperm proportion as indicated by iodine stain. Data pooled from independently crossed biological replicates and analysed by one-way ANOVA with multiple comparisons (Table 4.10), p value of <0.0001 indicated by ****.

In order to determine if the *Zm*MLP1-GFP was able to complement *ZmMlp1* RNAi at early developmental stages and recover the DEP phenotype, seeds from the reciprocal crosses were analysed at 6 DAP. Seeds were sampled, fixed (see 2.5.1), sectioned (see 2.5.2) and stained for histological analysis (see 2.5.8). Similarly to previous results, the *Zm*MLP1-GFP line was able to rescue the DEP phenotype, but only when de-

Table 4.9: One-way ANOVA multiple comparison test (Sidak) of proportional starch area from reciprocal and control crosses at 15 DAP. (RNAi = Line T02297, MLP1-GFP = Line T02492). n = number of kernels pooled from 3 biological replicates.

Cross Orientation	Mean starch proportion (%)	n	Mean Difference	95 % CI of MD	Significance
T02492 x T02297	50.03	116	-13.34	-16.31 to -10.36	****
A188 x T02297	36.69	116			
T02492 x T02297	50.03	116	-0.86	-4.83 to 3.12	ns
A188 x T02492	49.17	45			
T02297 x T02492	38.57	110	6.34	3.53 to 9.15	****
T02297 x A188	44.91	159			
T02297 x T02492	38.57	110	10.51	7.60 to 13.41	****
T02492 x A188	49.07	135			
A188 x T02492	49.17	45	10.60	6.60 to 14.61	****
T02297 x T02492	38.57	110			

Table 4.10: One-way ANOVA multiple comparison test (Sidak) of mature seed cross-sectional area from reciprocal and control crosses.(RNAi = Line T02297, MLP1-GFP = Line T02492). n = number of kernels pooled from 3 biological replicates.

Cross Orientation	Mean starch proportion (%)	n	Mean Difference	95 % CI of MD	Significance
T02492 x T02492	38.09	40	-0.999	-3.446 to 1.448	ns
T04292 x T02297	39.09	60			
T02297 x T02297	27.27	60	-1.003	-3.191 to 1.185	ns
T02297 x T02492	28.27	60			
T02492 x T02297	39.09	60	10.82	8.636 to 13.01	****
T02297 x T02492	28.27	60			

livered from the maternal line (Figure 4.11 (top)). The recovery of the DEP phenotype was then quantified by image analysis (see2.7.4), which showed statistically significant dimensional recovery of suspensor length and total embryo length only when *Zm*MLP1-GFP was the paternal parent plant (Figure 4.11 (a and d) and Tables 4.11 and 4.12).

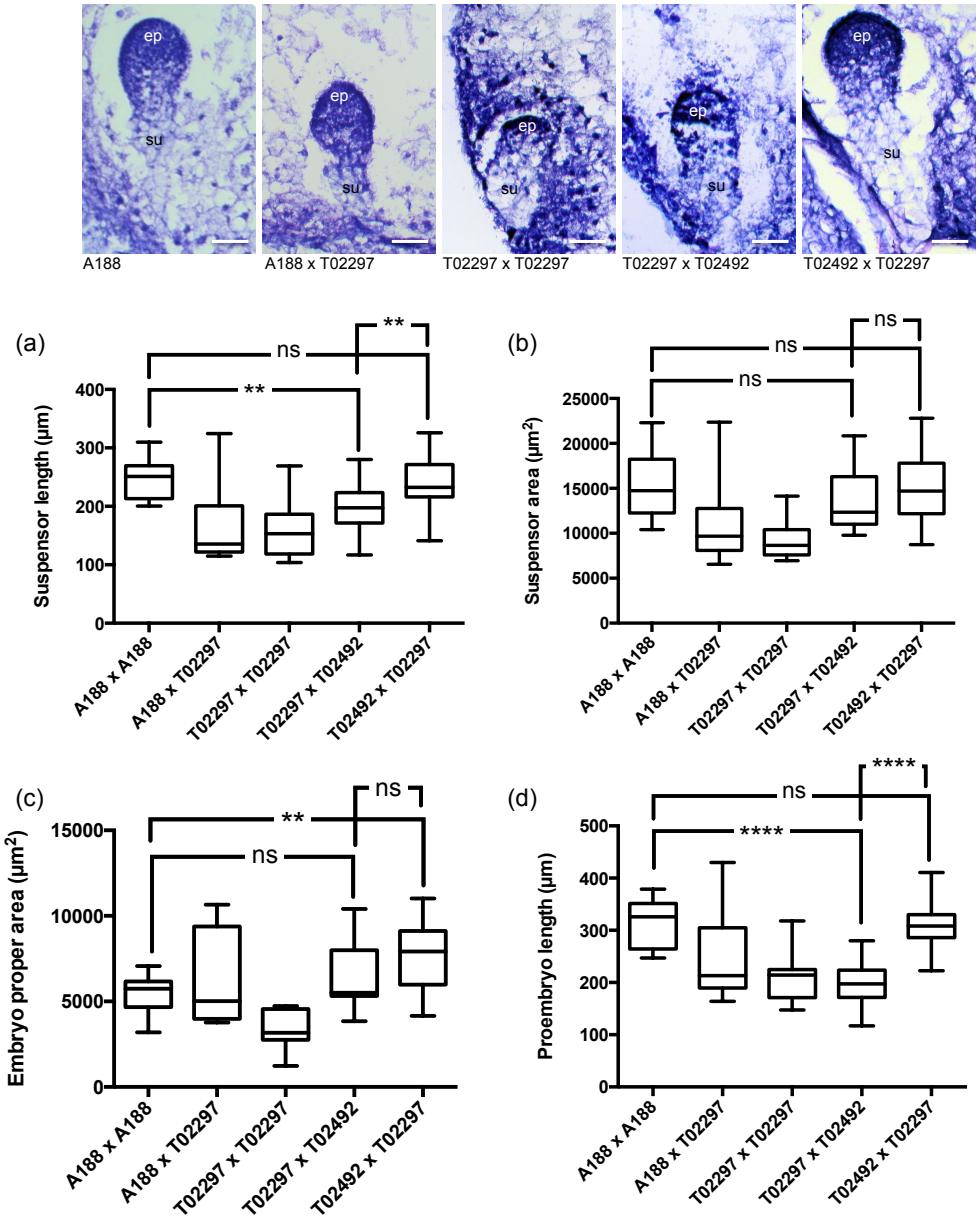

Figure 4.11: Histological analysis of genetic complementation of embryos derived from genetic complementation analysis of T02297 (*ZmMlp1* RNAi) crossed by T02492 (*Zm*MLP1-GFP). (Top) Embryo morphologies observed in reciprocal crosses between A188, T02297 and T02492 lines at 6DAP. (ep) = embryo proper, (su) = suspensor, scale bar = 50 µm. (a-d) Box and whisker plots of embryo dimensions measured in reciprocal crosses between A188, T02297 and T02492 lines at 6DAP. Data was analysed by one-way ANOVA with multiple comparisons (Table 4.11 and 4.12), p values of <0.01 and <0.0001 indicated by ** and **** respectively.

As a further measure, expression analysis was carried out by qRT-PCR to investigate the behaviour of *ZmMlp1*-GFP transcription at the molecular level. Reciprocal crosses were carried out between T02492 and A188 wild type lines (see 2.1.5), developing seeds sampled at 2, 4 and 6 DAP, RNA extracted (see 2.2.2), cDNA synthesised (see 2.2.3) and gene expression analysed by qRT-PCR (see 2.2.6). Synthetic *ZmMlp1*-GFP was shown to express at comparable levels between both genetic orientations at all three time points, indicating that expression of the transgene behaves similarly to the endogenous *ZmMlp1* gene, regardless of the parent of origin (Figure 4.12 (a)). The expression of the endogenous *ZmMlp1* and *ZmEsr6* genes was also found to be unaffected by the presence of the transgene in either genetic orientation aligning well with the wild type manifestation of the T02492 lines (Figure 4.12 (b and c)).

Table 4.11: One-way ANOVA multiple comparison test (Sidak) of embryo suspensor dimensions at 6 DAP from reciprocal and control crosses. (RNAi = Line T02297, MLP1-GFP = Line T02492). n = number of kernels pooled from 3 biological replicates.

MLP1 RNAi Line	**Mean suspensor length (µm)**	**n**	**Mean Difference**	**95 % CI of MD**	**Significance**
T02492 x T02297	241.4	23	-4.179	-39.04 to 30.69	ns
A188 x A188	245.5	13			
T02297 x T02492	196.3	16	-49.20	-86.72 to -11.68	**
A188 x A188	245.5	13			
T02492 x T02297	241.4	23	45.02	12.31 to 77.73	**
T02297 x T02492	196.3	16			
	Suspensor area (μm^2)				
T02492 x T02297	14790	23	-205.4	-3433 to 3022	ns
A188 x A188	15175	13			
T02297 x T02492	13633	16	-1543	-5015 to 1930	ns
A188 x A188	15715	13			
T02492 x T02297	14970	23	1337	-1690 to 4365	ns
T02297 x T02492	13633	16			

Table 4.12: One-way ANOVA multiple comparison test (Sidak) of embryo proper and total proembryo dimensions at 6 DAP from reciprocal and control crosses. (RNAi = Line T02297, MLP1-GFP = Line T02492). n = number of kernels pooled from 3 biological replicates.

MLP1 RNAi Line	**Mean embryo proper area (µm²)**	**n**	**Mean Difference**	**95 % CI of MD**	**Significance**
T02492 x T02297	7585	23	2164	624.4 to 3703	**
A188 x A188	5421	13			
T02297 x T02492	6619	16	1198	-458.5 to 2854	ns
A188 x A188	5421	13			
T02492 x T02297	7585	23	965.8	-478.3 to 2410	ns
T02297 x T02492	6619	16			
	Proembryo length (µm)				
T02492 x T02297	311.7	23	-2.272	-40.30 to 35.75	ns
A188 x A188	314.0	13			
T02297 x T02492	196.3	16	-117.7	-158.6 to -76.75	****
A188 x A188	314.0	13			
T02492 x T02297	311.7	23	115.4	79.73 to 151.1	****
T02297 x T02492	196.3	16			

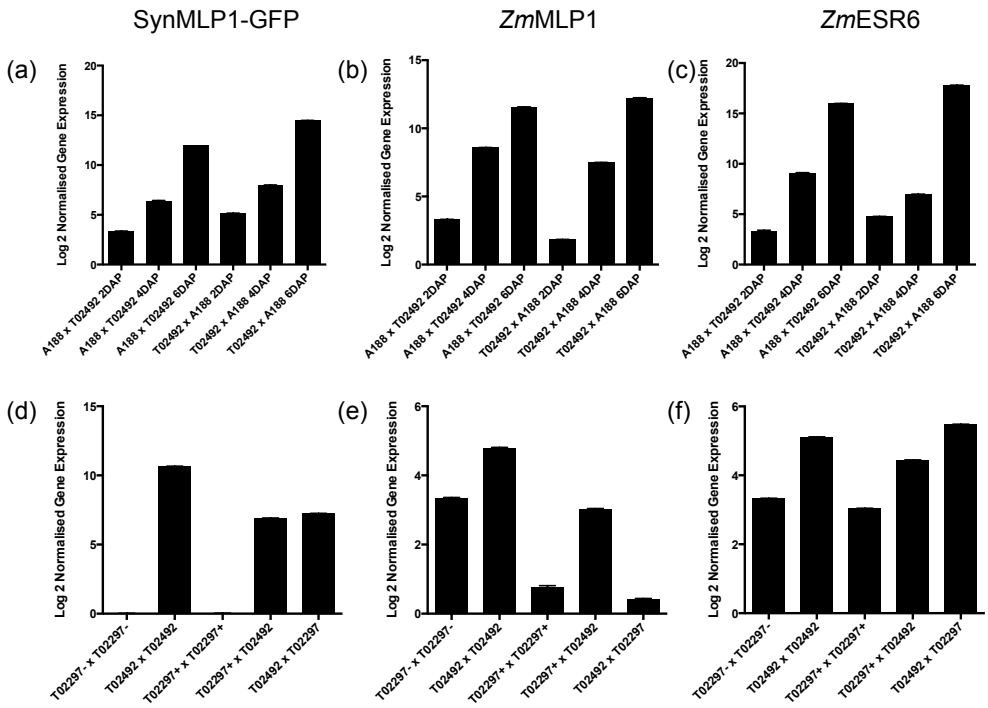

Figure 4.12: (a-c) Expression profile of reciprocal crosses between A188 and T02492 lines at different developmental stages by qRT-PCR. (d-f) Expression analysis of reciprocal crosses between T02297 and T02492 lines at 6 DAP by qRT PCR. Expression analysis of (a and d) synthetic *ZmMlp1*-GFP, (b and e) *ZmMlp1*, and (c and f) *ZmEsr6*. (+) Transgenic sibling plants; (-) Non-transgenic sibling plants. Error bars = standard error of the mean of technical replicates.

This analysis was then repeated using cDNA generated from seeds from reciprocal crosses between the *ZmMlp1* RNAi and *Zm*MLP1-GFP lines sampled at 6 DAP. Expression of the synthetic gene was detected at similar levels, in both orientations of the reciprocal crosses between T02492 and T02297 indicating no parent of origin effect (Figure 4.12 (d)). The expression of the synthetic gene in the T02492 sib crossed samples was found to be slightly higher, most-likely caused by a copy number effect (Figure 4.12(d)). Interestingly, the expression of the endogenous *ZmMlp1* transcript was found to be higher at 6DAP in when the RNAi was delivered from the female (Figure 4.12(e)), which could suggest a maternal effect, but contradicts somewhat the expression ana-

lysis observed in Figure 4.12 (a and b). However, the most likely explanation, is that similarly to the observations made in the initial RNAi line screening (Figure 4.1), there is variability between samples in the efficiency of the RNAi down-regulation. The slight variability in expression of *ZmEsr6* (Figure 4.12 (f)) also indicates that there could be some small variation in developmental stage of the sampled seeds, which could also contribute to higher expression of the endogenous *ZmMlp1* gene.

4.5 Discussion

The maize ESR is a clearly defined, specialised endosperm tissue type, with distinct morphological and genetic characteristics and a tightly regulated lifespan during early seed development. Furthermore, homologous tissue types can be found surrounding or adjacent to the embryo in other plant species such as *Arabidopsis thaliana* (Brown *et al.* , 2003; Mansfield & Briarty, 1990), wheat (Smart & O'brien, 1983) and barley (Brown *et al.* , 1994), suggesting a strong evolutionary conservation pressure and therefore a critical function. However, the exact function of the ESR has remained largely unknown (Cossegal *et al.* , 2007). In this chapter, results have been presented offering the first direct evidence for a function of the maize ESR through functional knock-down studies.

A crucial step in this functional investigation was the generation of multiple independently transformed transgenic lines, able to efficiently down-regulate the expression of *ZmMlp1* by RNA interference. These RNAi lines yielded seeds with a defective kernel phenotype similar to those observed in *ZmMeg1* RNAi (Costa *et al.* , 2012) and *miniature1* mutants (LeCLere *et al.* , 2010). The observed significant reduction in seed weight and size indicated a critical role for *Zm*MLP1 and hence the ESR, in kernel development or grain filling, though interestingly the proportion of the seed occupied by starchy endosperm did not appear to be affected. However, the reduction in grain

weight causes a reduction in absolute endosperm quantity and therefore most likely nutritional content of the seed, which could implicate *ZmMlp1* as an agronomically important gene.

Initially, it was hypothesised that the function of *Zm*MLP1 in the ESR could be analogous to that of *Zm*MEG1 in the BETL, playing a role in the normal development of the ESR tissue and its associated function. However, the *ZmMlp1* RNAi lines yielded seeds with developmentally and morphologically normal ESRs. Additionally, several ESR specific genes were observed to be unaffected by the down-regulation of *ZmMlp1*, which when combined with the localisation of the expressed *Zm*MLP1 protein to the surface of the embryo indicates that the function of *Zm*MLP1 may be in the embryo itself. Several BETL specific genes were also shown to be down-regulated which could be evidence for a signalling pathway between the developing embryo and the BETL.

As previously discussed, RNAi was the technology adopted for this research due to its established methodology, target specificity, and efficacy at down-regulation in maize. However, the best approach would probably be to use a complete knockout mutant for *ZmMlp1*, to ensure that there were no off-target effects influencing the phenotypic characterisation. After database searches, it was found that no mutant was available for *ZmMlp1*, leading to the RNAi based approach. However, recent advances in gene editing technologies could present an alternative strategy for creating a *ZmMlp1* knockout mutant. One such approach could be the use of transcription activator-like effector nucleases (TALENs), which use a repeat-rich protein sequence with sequence-specific DNA binding ability to target gene sequences for cleavage (Boch *et al.* , 2009). Another development has been in the development of the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated (CRISPR/Cas) system. This innovation is a two part system using a bacterial endonuclease Cas9, which can be targeted to specific sequence sites by engineering a separate chimeric guide RNA to containing a recognition

sequence (Mussolino & Cathomen, 2013). The Cas9 endonuclease then introduces a double stranded DNA break, with subsequent natural mutations arising in the sequence repair. Although both technologies are arguably still in their infancy, their application has been successfully applied to a variety of plant species (Puchta & Fauser, 2013). In fact, recent work has made use of both technologies to achieve gene editing on a range of target genes in maize (Liang *et al.* , 2014), demonstrating the suitability of this method for potential use in the work reported here.

The main effect of the achieved RNAi mediated *ZmMlp1* knock-down though, was the observed DEP phenotype in the developing embryo, showing *Zm*MLP1 to be involved in the normal pattern formation during early developmental stages. The DEP phenotype consisted of a morphologically altered pro-embryo with a marked reduction in the length of the suspensor and the dimensions of the embryo proper. One challenge in the characterisation of the DEP phenotype was identifying the cell types in the embryo in order to distinguish between the embryo proper and the suspensor. The literature contains very few genes that have been localised exclusively to one cell type in the maize embryo which limited the options for labelling. mRNA probes for *LTP1* (GRMZM2G101958), *LTP2* (GRMZM2G083725) and *OCL1* (Ingram *et al.* , 2000) were all tested as markers to differentiate the embryo proper from the suspensor but were found to be ineffective. The approach taken was to histologically stain the sections and subjectively differentiate the cell types based on observed cytological differences. To try and reduce any subjective bias in this method, the samples were analysed blind and identified after the measurements had been taken. An alternative approach could be to introgress the RNAi into a line expressing YFP tagged PIN1, an auxin transporter that is expressed in discrete regions of the embryo proper at early developmental stages (Chen *et al.* , 2014). This would also allow the investigation into whether auxin based signalling was affected in the *ZmMlp1* RNAi plants, potentially offering an explanation for the

observed phenotype.

This suspensor phenotype is analogous to the short suspensor phenotypes observed in *Arabidopsis* lines with loss of function mutations in genes of the *YODA* signalling pathway. *YODA* (*YDA*), *SHORT SUSPENSOR* (*SSP*) and *Embryo Surrounding Factor 1.3* (*ESF1.3*) have been shown to act synergistically to regulate the growth and development of the *Arabidopsis* embryo, resulting in suspensor abnormalities upon loss of function (Lukowitz *et al.* , 2004; Bayer *et al.* , 2009; Costa *et al.* , 2014). Interestingly, *At*ESF1.3 bears striking homology with *Zm*MLP1 at the amino acid level, perhaps indicating an, as yet, undiscovered *YODA*-like pathway in maize, in which *Zm*MLP1 could play an interacting role.

Since *Zm*MLP1 is expressed exclusively in the ESR and is required for normal embryo pattern formation, we can conclude that one role of the ESR is to interact with the embryo and ensure its normal development. The global endosperm expression analysis of the *ZmMlp1* RNAi lines as performed by RNAseq, indicated up-regulated genes enriched in transcriptional regulators and signalling peptides, and a corresponding down-regulation in CRPs and metabolic regulators. However, the fold change effects observed in the RNAseq experiments are relatively low, probably due to the use of segregating ears during sampling, effectively diluting the effect of the RNAi construct. Ideally, the experiment would be repeated using homozygous RNAi plants, however, it was observed that homozygous plants from the RNAi lines produced very unhealthy ears with very deformed seeds. This may have introduced interfering results due to the presumed highly disturbed transcriptional status of the plants. The observations though, support previous hypotheses that the ESR functions as an interfacing tissue for cross-talk between the embryo and endosperm. Furthermore, the proposed role of the ESR in embryonic development is compounded by the association of the discussed CLAVATA1-like and Subtilisin-like genes, which are up-regulated in response to *Zm*MLP1 knock-down.

A hypothesis could be that these genes are involved in the same pathway as *Zm*MLP1, and act in concert to regulate some aspect of embryonic development.

The overall defective kernel phenotype could be explained by two-way signalling between the embryo and endosperm. When the embryo is developmentally altered, this could alter its demand for sugars and nutrients, resulting in a reduced sink tissue. In turn, a build-up of these factors in the endosperm could cause the loss of nutrient balance in the developing seed and reduced grain filling as a consequence. Sugars have been implicated in important signalling roles when associated with cell wall invertase activity, potentially regulating cell differentiation and division in early seed development through the establishment of glucose spatial gradients (Bate *et al.* , 2004; Nowack *et al.* , 2007; Wang & Ruan, 2012b). The observed expression augmentation of sugar transporters and invertase inhibitors could lead to a loss of sugar associated signalling homeostasis and a resulting alteration in pattern development in the embryo. Additionally, the observed up-regulation of the *Zm*SH-1 sucrose synthase could be responsible for a development of a *Shrunken1*-like phenotype, which could also be linked to the *Zm*MLP1 pathway.

Another interesting result came from the genetic complementation of *ZmMlp1* RNAi, showing that the phenotype could only be rescued when the RNAi insensitive synthetic *ZmMlp1* transcript was delivered through the maternal line. This suggests that *Zm*MLP1 is functionally critical either before fertilisation in the central cell, or in the very early stages of endosperm development immediately after fertilisation. This could be due to a transcriptional delay in the expression of *Zm*MLP1 from the paternal gamete and hence, a deleterious effect on early embryo pattern establishment. Additionally, the expression profile in both cross-orientations of the synthetic *ZmMlp1* gene after fertilisation was observed to behave similarly to the endogenous gene. This suggests that while the maternal copy of the gene is clearly important due to due to the absence

of paternal expression before and immediately after fertilisation, there appears to be no maternal effect at work in this gene, unlike the homologue *ZmMeg1* (Gutierrez-Marcos *et al.* , 2004; Costa *et al.* , 2012). The successful recovery of the phenotype by *Zm*MLP1-GFP also indicates that the GFP tagged version of the protein is functional, despite the relative size of GFP compared to *Zm*MLP1. This allows greater confidence in the GFP localisation studies carried out in chapter 3 and also presents the *Zm*MLP1-GFP as a potentially useful tool for future work in identifying interacting partners through co-immunoprecipitation and proteomics experiments. Having identified a role for *Zm*MLP1 in maize development, the next step required is to investigate the structural properties of *Zm*MLP1 and understand how these are related to the function.

5 Structure-Function Analysis of *Zm*MLP1

5.1 Introduction

*Zm*MLP1 has been shown in previous chapters to exhibit a unique and tightly controlled expression pattern in maize seeds and has an important role in seed development. At this point, an interesting biological question presents itself: How does *Zm*MLP1 elicit this function with regards to its protein structure? In order to address this question, the structural properties of *Zm*MLP1 need to be investigated. As previously discussed, *Zm*MLP1 is part of the small CRP family of proteins, characterised by their low molecular weight, predicted signal peptide and conserved cysteine rich domain.

Plant CRPs make up a highly abundant class of genes, in some cases representing as much as 3 % of the observed gene compendium (Silverstein *et al.* , 2007). Given the homology in this gene class, and the wide range of attributed functions as discussed in chapter 1, it is likely that evolution has positively selected the CRP general structure as a multi-functional platform on which to build unique functional molecules. As with *Zm*MLP1, it could also be that tightly controlled expression patterns govern a large part of the function and that the protein molecules could be fairly generic in nature.

Few CRPs have been structurally investigated thus far, some exceptions including *At*EPFL9 (Ohki *et al.* , 2011), *Br*SP11 (Mishima *et al.* , 2003) and *At*ESF1.3 (Costa *et al.* , 2014), for several reasons. The CRP class was only discovered relatively recently and there are several major challenges in working with these proteins. The foremost problem in structural studies of CRPs has historically been the lack of an efficient expression system capable of producing correctly folded proteins due to the presence of multiple disulphide bonds. Furthermore, the non-sequential nature of the disulphide bonds and the lack of example structures presents a major challenge in the alternative approach of *in vitro* refolding of incorrectly folded proteins.

There are many expression systems available for recombinant protein production us-

ing both bacterial and plant based expression systems. *E.coli* is a well researched and relatively reliable expression system, and offers the advantages of inexpensive media, it produces large volumes of high density cell cultures due its rapid growth (Vallejo & Rinas, 2004) and a wide variety of available expression vectors. However, due to the eukaryotic origins of *Zm*MLP1, and perhaps more crucially, the 4 predicted disulphide bonds and corresponding complex folding, *E.coli* may not be able to produce an active form of the protein, due to a lack of post-translational modification and correct folding machinery (Nguyen *et al.* , 2011). Specifically, *E.coli* lacks the ability to form disulphide bonds in its cytoplasm due to a reducing environment, which tends to lead to the formation of insoluble aggregates known as inclusion bodies, though these can potentially be recovered and re-folded to an active form (Baneyx & Mujacic, 2004; Vallejo & Rinas, 2004). A potential solution to this problem is to use engineered expression strains of *E.coli* that are modified to create a more oxidising cytoplasm through the augmentation of redox pathways (Nguyen *et al.* , 2011; Lobstein *et al.* , 2012; Saaranen & Ruddock, 2012; Hatahet *et al.* , 2010). Alternatively, the expression of the protein could be targeted to the periplasm of *E.coli* where disulphide bonds can be readily formed, however, the protein yield in such systems is usually low (Denoncin & Collet, 2013; de Marco, 2009).

The investigation of protein structure is a large field, with over 10,000 protein structure files deposited in the protein data bank (PDB) to date. There are three main approaches to resolving the 3D structure of a protein: X-ray crystallography, NMR spectroscopy and cryo-electron tomography. The vast majority of protein structures in the PDB have been determined by x-ray crystallography, a technique involving the crystallisation of the protein followed by x-ray bombardment and interpretation of the resulting x-ray diffraction pattern. This technique has the greatest precision with atomic resolution achievable on large proteins and complexes. However, many proteins can be extremely

difficult or even impossible to crystallise to a high enough quality for x-ray experiments due to intrinsic structural disorder or lack of a highly homogenous sample (Feng *et al.* , 2011). Furthermore, interpretation of the data requires a great deal of expertise and the whole process can be very slow. Cryo-electron tomography is a technique that requires immobilising the hydrated protein in a thin layer of ice, followed by a series of electron microscopy imaging experiments performed at different tilt angles in order to gain 3D projection data. This technique is particularly useful in the structural determination of large complexes, as used in the investigation of the clathrin cage (Young, 2007), but is less useful for small proteins.

Structural NMR is a growing technique with applications in both solid and solution state samples, contributing nearly 10 % of the resolved structures in the PDB. Although lacking in resolution compared to x-ray crystallography, NMR has several advantages, the largest being the lack of requirement for protein crystallisation, needing only a protein solution, at concentrations as little as 10 µM. Furthermore, in solution state, NMR is a rapid technique, requiring only a few hours, is able to detect solution/ligand induced conformational changes and generate structures from inherently disordered protein regions (Feng *et al.* , 2011).

The work reported in this chapter utilises both bacterial and plant based expression systems to produce recombinant *Zm*MLP1 in order to assess their suitability for structural investigation. Preliminary structural work is then carried out by enumeration of disulphide bonds by means of biochemical assay for free thiol groups. Secondary structure of the recombinant protein is investigated using circular dichroism, a technique relying on the differential absorption of left and right circularly polarised light by secondary structural elements (Kelly *et al.* , 2005).

Investigations were also carried out with collaborators in Japan to solve the 3D structure of *Zm*MLP1. This was carried out by use of a plant cell culture based, inducible

virus-mediated expression system, allowing high yield expression of plant proteins in their native structural conformations (Dohi *et al.* , 2006; Ohki *et al.* , 2008). This is followed by investigation into the 3D structure by solution state NMR for the reasons discussed above. Experiments are then reported that probe the structure-function characteristics of the CRP family by generating sequence chimeras with a characterised protein *At*ESF1.3 and expressing them in transformed *Arabidopsis* lines.

Arabidopsis however, exhibits a rather different reproductive developmental pattern than that of maize which is discussed in chapter 3. As a dicot, *Arabidopsis* seeds develop a transitory endosperm around the developing embryo which serves to supply the embryo with nutrients during early developmental stages. Unlike in maize, this endosperm is fully reabsorbed upon embryo maturity. However, a region of the micropylar endosperm that completely envelops the early embryo and is later restricted to surrounding the basal part of the suspensor seems to be functionally similar to the ESR of maize, as demonstrated by the characterisation of *At*ESF1.3 (Costa *et al.* , 2014). Figure 5.1 illustrates the different developmental pattern adopted by *Arabidopsis* compared to maize (see chapter 3). One of the notable differences that is particularly relevant to this work is the rapid elongation of the suspensor after the first asymmetric zygotic division. This elongation process is regulated by *At*ESF1.3 (Costa *et al.* , 2014) which when knocked down, produces short suspensors, while the opposite is achieved upon over-expression. In this chapter, this regulatory effect is utilised as a tool to investigate the functional properties of CRPs across species in the pre-globular stages of development (Figure 5.1).

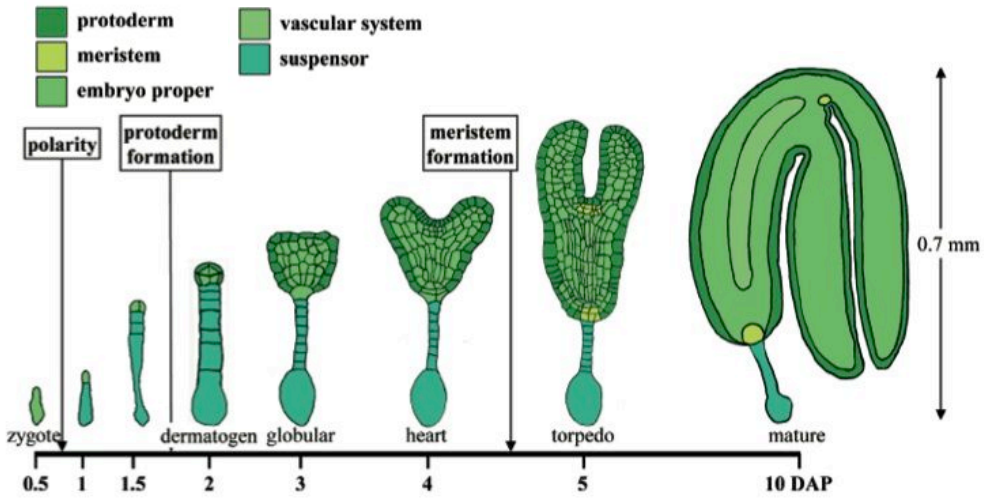

Figure 5.1: Schematic representation of the developmental stages of the *Arabidopsis* embryo. Different cell types are indicated by colour scheme and the approximate developmental stage is annotated by the number of days after pollination (DAP). Adapted from Hajduch *et al.* (2005).

The aims of this chapter are to investigate the technical challenges of structure based analysis of CRPs, structurally resolve *Zm*MLP1 and explore the relationship between the structural features and the established function of *Zm*MLP1. This data will then be discussed and contextualised with regards to CRPs in varying functional roles and across different species.

Results

5.2 *In silico* structural analysis of *Zm*MLP1

As a preliminary analysis, the *Zm*MLP1 sequence was investigated using the PredictProtein web server (Rost *et al.* , 2004) to generate a predicted secondary protein structure. The 8 cysteine residues were predicted to form 4 non-sequential disulphide bonds (Figure 5.2 (b)) which suggests complex protein folding, requiring oxidation and isomer-

isation steps in order to form the correct cysteine pairs (Hatahet *et al.* , 2010; Nguyen *et al.* , 2011). Furthermore, the prediction suggested that there may be two helical segments, separated by a beta sheet, incorporated in the compact disulphide bond-rich region in the core of the sequence, with a greater level of disorder predicted in the N and C terminal regions (Figure 5.2 (b)). Predicted interaction regions were observed in both the variable region of the sequence, and on the loops generated by the disulphide bonds in the CRD. This could confer regions of specific activity as found by Ohki *et al* in the similar stomagen protein (Ohki *et al.* , 2011).

5.3 Expression of active *Zm*MLP1

In order to validate the *in silico* data, sufficient *Zm*MLP1 was required to perform direct structural studies. As discussed in chapter 3, the spatial expression region of *Zm*MLP1 is restricted to a very small region of the seed, making isolation of the native protein impractical for this purpose. Therefore, a dual-platform approach was taken to obtain recombinant *Zm*MLP1 protein, using both bacterial and plant based expression systems.

To determine if *E.coli* was suitable for expression of correctly folded *Zm*MLP1, His-tagged synthetic *Zm*MLP1 was produced in an expression strain (see 2.4.1), purified by immobilised metal affinity chromatography (IMAC) (see 2.4.9) and its folding analysed. Using this technique, approximately 25 mg of protein was purified from 2 litres of culture to a final concentration of 5.34 mg mL^{-1}. The Ellman's reaction was performed to assess the degree of folding by quantification of formed disulphide bonds (see 2.4.12). This technique uses reduced and oxidised forms of the protein sample in a colorimetric assay to quantify the number of free thiol groups compared to a known standard of reduced glutathione. This experiment yielded a ratio of reduced to non-reduced thiol groups

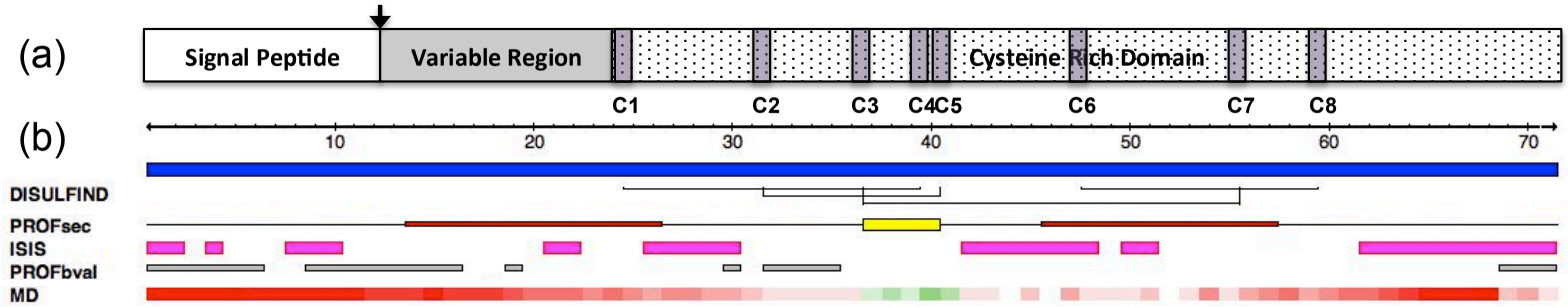

Figure 5.2: General and predicted secondary structure of *Zm*MLP1. (a) A schematic representation of the general CRP structure found in *Zm*MLP1. Arrow indicates the predicted cleavage site of the signal peptide. (b) *In silico* secondary structural prediction obtained by amino acid sequence submission to the online PredictProtein server. DISULFIND indicates predicted disulphide bonds (Ceroni *et al.* , 1), PROFsec indicates alpha helices and beta sheets shown in red and yellow respectively (Rost *et al.* , 2004), ISIS predicts regions of protein-protein interaction (Ofran & Rost, 2007), PROFbval indicates regions with a high degree of predicted flexibility (Schlessinger *et al.* , 2006; Schlessinger & Rost, 2005) and MD predicts regions of order (green) and disorder (red) within the protein structure (Schlessinger *et al.* , 2009). Scale bar = amino acid residues.

of ~4, perhaps indicating the presence of only one disulphide bond in the *Zm*MLP1 sample most-likely due to the expected mis-folding of the protein in the *E.coli* expression system. The result however, could also in part be affected by the presence of multiple species in the sample. Though the SDS-PAGE analysis showed the protein to be pure, the chromatogram in chapter 3 (Figure 3.7 (c)), shows the elution of *Zm*MLP1 in 3 separate phases. This could be due to differential affinity for the nickel column of several different conformational forms of the protein.

As a further test, to assess whether any of the protein secondary structure resembled the predicted data, the purified *Zm*MLP1 sample was subjected to circular dichroism (see 2.4.13). Circular dichroism is a structural analytical technique which exploits a molecule's differential absorption of left and right circularly polarised light, in accordance with differing secondary structural features. This leads to absorption peaks at characteristic wavelengths, allowing the secondary structure to be calculated (Kelly *et al.* , 2005). This data was then submitted to the Dichroweb server (Lobley *et al.* , 2002; Whitmore & Wallace, 2008) and analysed using the CDSSTR algorithm (Sreerama & Woody, 2000) and the SP175 reference database (Lees *et al.* , 2006). A secondary structure for *Zm*MLP1 not too dissimilar from the predicted form (see 5.2) was elucidated (Figure 5.3 and Table 5.1). The normalised root-mean-square deviation (NRMSD) value is 0, normally indicating a good fit, however, this structure calculation seems questionable when the plot is analysed manually. The fit plot lacks minima at 208 and 222 nm, characteristic of alpha helix structure (Kelly *et al.* , 2005) even though the algorithm has interpreted a ~60% helical structure. Further suspicion stems from the intense minimum peak seen at 200 nm, a feature typically seen in disordered proteins (Mouillon *et al.* , 2006; Uversky, 2002). Combined with the incomplete formation of disulphide bonds (see 2.4.12), the data suggests that the DichroWeb interpretation is anomalous, and the protein is incompletely folded.

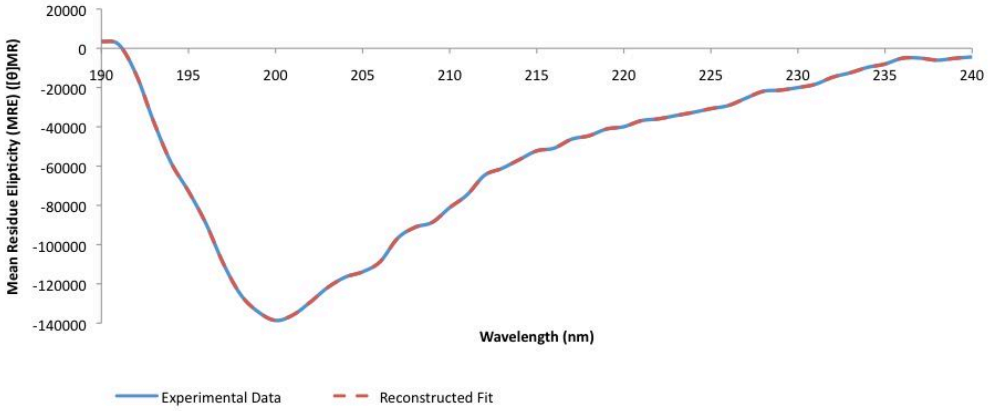

Figure 5.3: Circular Dichroism spectrum fit for recombinant *Zm*MLP1. Analysis was performed using the DichroWeb server, implementing the CDSSTR method and the SP175 reference database.

A second strategy involves use of a *Zea mays* plant cell culture. This offers the major advantage of native folding pathways, ensuring the production of an active form of the protein. However, some disadvantages of this system include: lengthy culture set up, slower growth rate, lower cell density along with significant problems in protein aggregation and degradation, collectively leading to a lower yield potential (Doran, 2006; Shih & Doran, 2009; Hellwig *et al.* , 2004). In order to test the viability of this system, de-differentiated cell cultures were established using existing dexamethazone (DEX)-inducible over-expression *ZmMeg1*/pro*ZmMeg1*-GUS maize lines (Costa *et al.* , 2012). Transgenic seeds were germinated and root explants used to generate calli on solid media (see 2.8.1). Calli were then genotyped (see 2.2.7 and 2.2.8) to select transgenic and wild type calli, and a selection of positive and negative tissue lines maintained. Calli were then induced with DEX (see 2.8.3) and assayed for GUS activity to assess promotor activity (see 2.4.14). Promotor activity was observed in both the induced and un-induced transgenic lines as indicated by GUS staining, but no expression was observed in the wild type samples, indicating a leaky vector. Liquid cultures were then generated (see 2.8.2), induced (2.8.3), and the protein extracted from the media by

means of acetone precipitation (see 2.4.4). The protein samples were then analysed by SDS-PAGE (see 2.4.6) and Western blotting (see 2.4.15) using a *Zm*MEG1 antibody. Despite the observed promoter activity, no *Zm*MEG1 protein was detected, which was probably due to the aforementioned difficulties with yield potential.

Given the difficulties encountered with the maize cell culture, a synthetic gene construct was designed for the expression of *Zm*MLP1 using a transient expression system in *Nicotiana benthamiana* leaves. The construct was designed to include a removable His-tag at the c-terminus by inclusion of a proteolytic cleavage site between the *Zm*MLP1 sequence and the tag (see 7.5). This gene was then cloned into the binary expression vector pEarleyGate100 (see 2.3.7 and 7.8), and transformed into *Agrobacterium* (see 2.1.10). In order to enhance the expression of *Zm*MLP1 in tobacco, the synthetic gene was co-expressed with the P19 protein in *Nicotiana benthamiana* leaves (see 2.1.11). P19 is a 19 kDa protein originating from the tomato bushy stunt virus which has been shown to reduce the effects of post-transcriptional gene silencing. This occurs by P19 sequestering the siRNA molecules involved in the recruitment of the RISC complex, leading to increased transgenic expression levels (Lakatos *et al.* , 2004; Voinnet *et al.* , 2003). The transformed leaves were then harvested, the protein extracted (see 2.4.4) and analysed by SDS-PAGE and Western blot (2.4.6 and 2.4.15). Significantly more *Zm*MLP1 protein was detected in the P19 co-expressed sample compared to the synthetic gene alone (Figure 5.4). Having successfully expressed *Zm*MLP1 in tobacco, the experiment was repeated on a larger scale and purification attempted by use of the His-tag. However, no protein was purified in this manner, suggesting that perhaps the yield was not sufficient to facilitate efficient affinity based binding. Although this purification strategy had proved ineffective, the expression capability in tobacco cells had been shown prompting a further investigation into alternative methods of production and isolation of *Zm*MLP1 using tobacco.

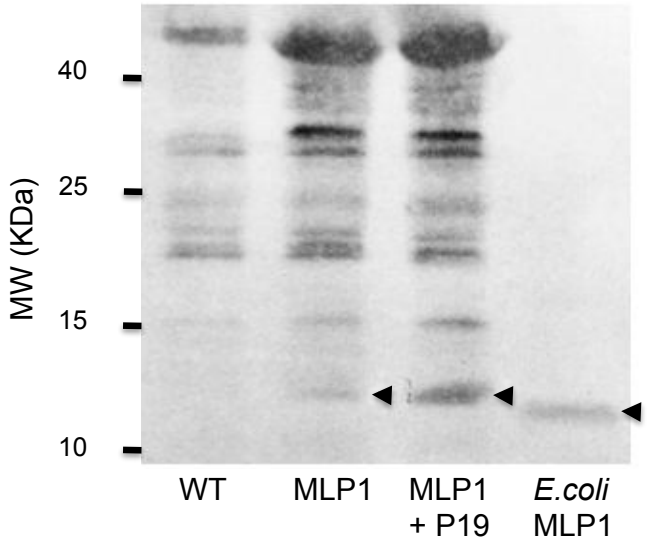

Figure 5.4: Expression of *Zm*MLP1 in tobacco. Western blot analysis of total protein extracted from tobacco leaves 2 days post infection. (i) WT plant, (ii) plants transformed with *Zm*MLP1 construct (see 7.5), (iii) plants co-transformed with *Zm*MLP1 and P19 constructs (Voinnet *et al.* , 2003), (iv) *E.coli* produced *Zm*MLP1 positive control. Arrows indicate *Zm*MLP1 detection.

5.4 Solving the 3D structure of *Zm*MLP1

Having established that tobacco cells were able to express *Zm*MLP1, a collaboration was launched with the groups of Professor Shinya Ohki at the Center for Nano Materials and Technology (CNMT), Japan Advanced Institute of Science and Technology (JAIST), and Professor Masashi Mori in the Laboratory of Plant Gene Technology, Research Institute for Bioresources and Biotechnology, Ishikawa Prefectural University. Professors Ohki and Mori have developed an inducible virus-mediated expression system using liquid BY-2 tobacco cell culture, enabling high yield expression of disulphide bonded plant proteins in their native structural conformations (Dohi *et al.* , 2006; Ohki *et al.* , 2008). This expression system makes use of the LexA-VP-16-ER (XVE) transactivation mechanism, which in the presence of estradiol, activates the expression of

a replicating viral vector containing the foreign gene construct in place of the viral coat protein ORF. This causes a build up of viral RNA in the BY-2 cells, leading to efficient production of foreign proteins. When combined with the use of stable isotope-labelled nitrogen sources in the BY-2 growth media, uniform protein isotope-labelling can be achieved, allowing structural analysis by solution state nuclear magnetic resonance (NMR) spectroscopy (Ohki *et al.* , 2008; Dohi *et al.* , 2006; Ohki *et al.* , 2011; Costa *et al.* , 2014).

To apply this technique to *Zm*MLP1, a synthetic gene construct was designed with the full amino acid sequence of *Zm*MLP1 followed by a protease cleavage site and a His-tag for purification (see 7.5). This construct was then cloned by our collaborators into the pBICER8-ToMV plasmid and co-transformed with a Ti plasmid, which constitutively expresses XVE, into BY-2 cells using *Agrobacterium* (Dohi & Mori, 2007; Dohi *et al.* , 2008, 2006). Similar to the methodology used for STOMAGEN and *At*ESF1.3 (Ohki *et al.* , 2011; Costa *et al.* , 2014), tBY-2 cells were grown in ^{15}N media to isotopically label the proteins, expression was induced with estradiol and the secreted *Zm*MLP1 protein purified from the media by nickel affinity chromatography. The His-tag was then cleaved and the resulting protein was then further purified by HPLC in preparation for NMR analysis.

*Zm*MLP1 was then used in a Heteronuclear Single Quantum Coherence (HSQC) NMR experiment which yielded a ^{1}H-^{15}N HSQC spectrum showing the ^{1}H and ^{15}N chemical shifts of each residue (Figure 5.5 (b)). This identified the mature form of *Zm*MLP1 to be 64 amino acids in length, with the 8 cysteine residues coupled by 4 non-sequential disulphide bonds and featuring secondary structure comprising of three beta sheets and one alpha helix (Figure 5.5 (a)). These data verify the predicted cleavage of the signal peptide as discussed in chapter 3 and vastly improve the structural information compared to the *in silico* data in Figure 5.2. The NMR data was then used to construct

a 3D protein backbone structure, overlaying multiple structural maps for improved reliability (Figure 5.5 (c)). This 3D data illustrates the relatively rigid structure of the central core of the protein, contrasting with the flexible and disordered C and N terminal regions and various loop regions as indicated by the lack of consensus structure in these sequence sectors.

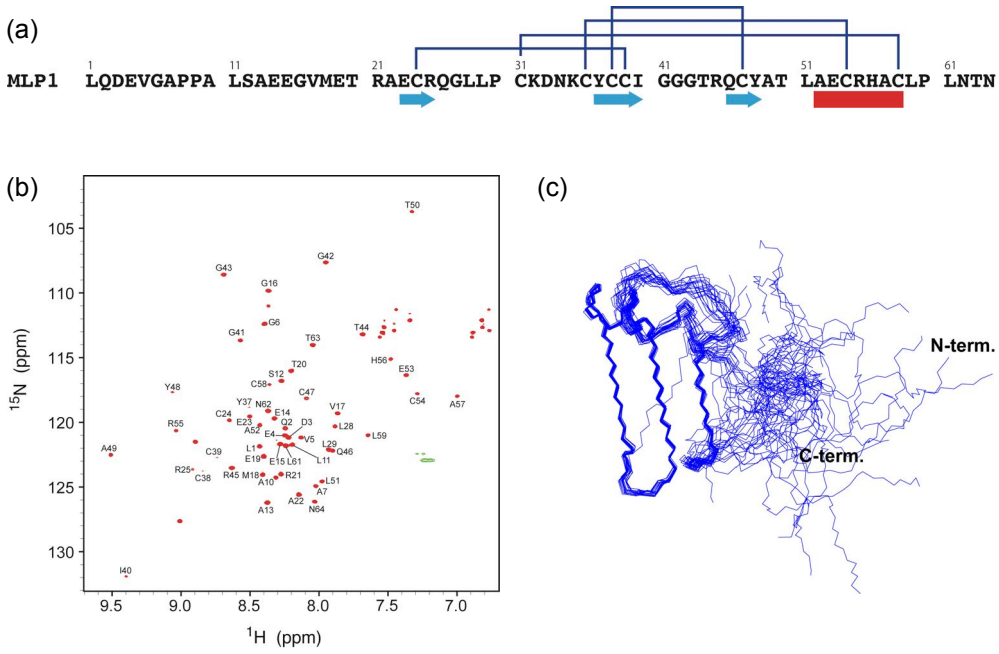

Figure 5.5: NMR structural conformation analysis of *Zm*MLP1. (a) Amino acid sequence of *Zm*MLP1 with annotated disulphide bonds (blue lines), alpha helix (red cylinder) and beta sheets (blue arrows) as resolved by structural NMR. (b)^{1}H-^{15}N HSQC of *Zm*MLP1 annotated with residue type and number. (c) Overlay of 20 *Zm*MLP1 NMR structures, displayed as protein backbones.

The NMR data was then used to generate 3D structural models, illustrating the relative positions of the secondary structure, disulphide bonds, disordered loop regions and the electrostatic charge distribution on the molecular surface (Figure 5.6 (a and b)). The ribbon model illustrates a disulphide bonded scaffold core, supporting several flexible loop regions, similar to the structural formats of *At*ESF1.3 and *At*EPFL9 (Costa *et al.* , 2014; Ohki *et al.* , 2011). Interestingly, the conserved aromatic residues, Y_{37} and

Y_{48} form a structural array very similar to the array conformation of the conserved tryptophan residues T_{35} and T_{48}in *At*ESF1.3, suggesting perhaps some functional importance. Furthermore, the core motif of *Zm*MLP1 shares similarities with the proteins of the cysteine knot (Knottin) family, many of which are involved in toxic and inhibitory functions (Pallaghy *et al.* , 1994; Norton & Pallaghy, 1998). The molecular surface model demonstrates the hydrophobic edge formed by the aromatic side chains, adjacent to a negatively charged pocket, which could be indicative of a functional binding site. This hypothetical binding site is placed under further suspicion by the presence of a predicted phosphorylation site on Y_{37} as predicted by the NetPhos 2.0 web server (Blom *et al.* , 1999), potentially implicating this structural feature as a target of kinase activity (Table 5.2). Furthermore, NetPhos 2.0 also predicted a phosphorylation site on S_{12}, a serine residue located in the flexible N-terminal region which is potentially long and mobile enough to reach the hypothetical binding site, allowing its involvement in molecular interactions (Table 5.2). Alternatively, if *Zm*MLP1 forms part of a multi-molecular complex, this second phosphorylation site could interact with a different kinase domain. Further experiments using the functional GFP tagged *Zm*MLP1 as described in chapters 3 and 4 are necessary to identify potential interacting proteins through mass spectrometry based proteomics and to investigate this phosphorylation hypothesis, since phosphorylation predictions are not particularly reliable.

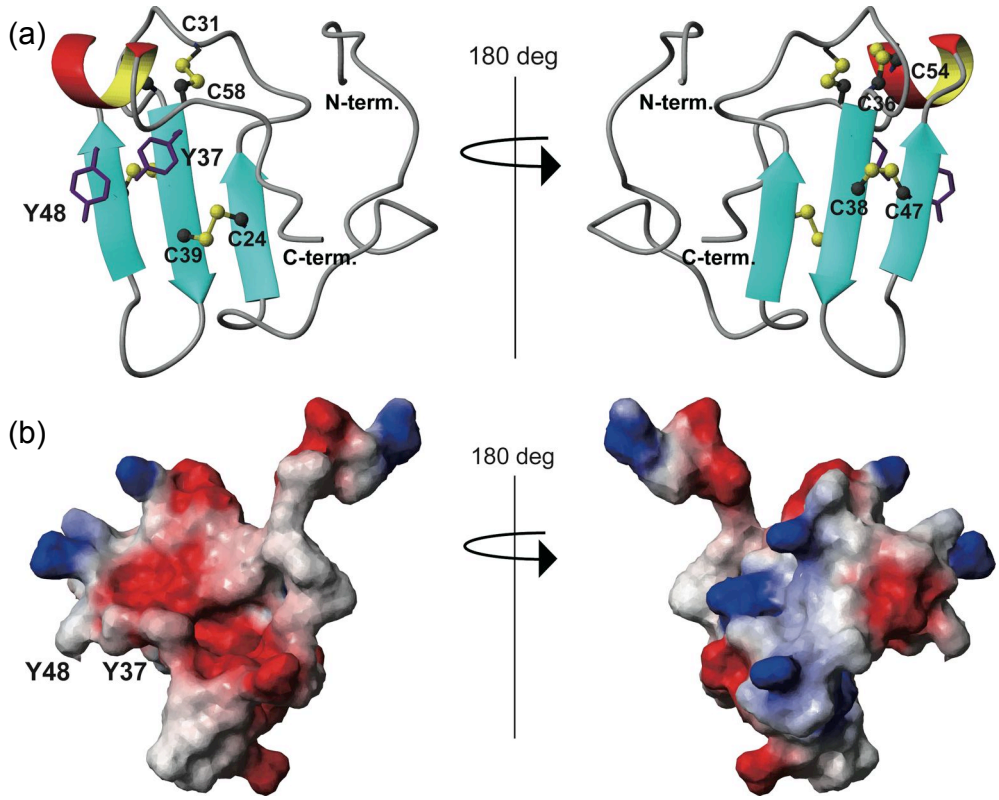

Figure 5.6: (a) *Zm*MLP1 NMR structure represented as a ribbon model. The location of the disulphide bonds formed by the cysteine pairs are shown, together with two conserved aromatic residues. (b) Surface charge model of *Zm*MLP1 showing electrostatic potential distribution. (red) Negatively charged , (blue) positively charged , and (grey) hydrophobic residues.

Table 5.1: Predicted secondary structure fractions calculated from experimental data submitted to the DichroWeb server using the CDSSTR method and the SP175 reference database. Normalised root-mean-square deviation (NRMSD) = 0

Structural feature	**Helix 1**	**Helix 2**	**Strand 1**	**Strand 2**	**Turns**	**Unordered**	**Total**
Fraction	0.46	0.14	0.03	0.05	0.10	0.21	0.99

Table 5.2: Predicted phosphorylation sites in the MLP1 amino acid sequence as predicted by the NetPhos 2.0 web based server (Blom *et al.* , 1999)

Target residue	**AA Position**	**Context**	**Score**	**Phosphorylation?**
Serine	12	PPAL**S**AEEG	0.995	Yes
Threonine	20	GVME**T**RAEC	0.141	No
	44	IGGG**T**RQCY	0.052	No
	50	QCYA**T**LAEC	0.054	No
	63	LPLN**T**N	0.019	No
Tyrosine	37	DNKC**Y**CCIG	0.677	Yes
	48	TRQC**Y**ATLA	0.138	No

5.5 Interspecific structure-functional analysis of CRPs

As mentioned in previous chapters, the amino acid structure of *Zm*MLP1 is characteristic of the plant CRP family, the general conserved structure of which is illustrated in Figure 5.2 (a). Focussing on *Zm*MLP1, several homologous proteins can be found across four different plant species identified by homology searches. *Zm*MLP1-like proteins have been identified in *Triticum aestivum* (wheat), *Hordeum vulgare* (barley) and *Arabidopsis thaliana* (Figure 5.7). All the homologues contain 8 cysteines in a similar configuration, predicted signal peptides and conserved aromatic residues located adjacent to cysteine residues. With the exception of *Zm*MEG1 (Gutierrez-Marcos *et al.* , 2004; Costa *et al.* , 2012) and *At*ESF1.3 (Costa *et al.* , 2014), the function of the *Zm*MLP1-like proteins is as yet unknown, but the level of conservation across many plant species suggests a functionally important role.

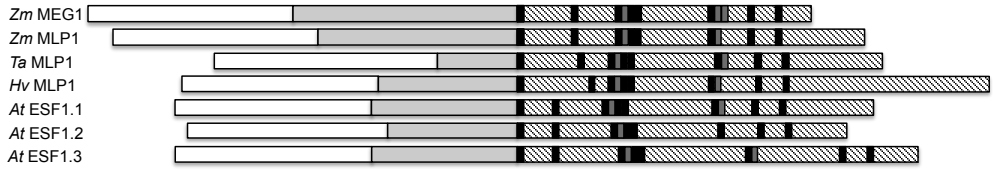

Figure 5.7: Schematic alignment of *Zm*MLP1 homologues across species indicating signal peptides (white), variable regions (grey) and cysteine rich domains (CRD) (hatching). Cysteine residues are indicated in black, while conserved aromatic residues in the CRD are indicated in dark grey.

The conserved primary structure configuration of the CRPs suggests a similar 3D structural format. As demonstrated by the solution state NMR structures of stomagen (*At*EPFL9) (Ohki *et al.* , 2011), *Br*SP11 (Mishima *et al.* , 2003) *At*ESF1.3 (Costa *et al.* , 2014) and now *Zm*MLP1, we can see that the cysteine rich domain (CRD) forms a core protein scaffold, by the formation of non-sequential disulphide bonds. The sequence segments between these S-S bonds form loops extending away from the core, some of which are involved in secondary structure. This core scaffold format allows for

an enormous degree of variability in the final structure and solution exposure of the inter-disulphide loops, owing to the high number of permutations available with regards to cysteine pairing. For example, a given protein sequence containing 8 cysteines, assuming they are all involved in S-S bonds, offers 2520 bond configurations, though this would likely be reduced through steric effects. When you then consider the variability in loop length and composition, the structural adaptability of this protein class is phenomenal.

Given the 3D structure of stomagen, further experiments confirmed that the core S-S structure is vital to the function of the protein (Ohki *et al.* , 2011). The loss of any of the stomagen S-S bonds by cysteine pair deletion, resulted in a non-functional, structurally disrupted protein which had formed S-S bonds incorrectly or not at all. Loss of function was also reported in scaffold/loop swapped and point mutated stomagen, identifying functionally important residues situated within one of the loops. The activity of this loop was maintained even when a different scaffold (*At*EPF3) was introduced, collectively demonstrating that both a CRD core and the functional loops must be intact to confer function (Ohki *et al.* , 2011). The extent to which the scaffold can be altered though, will likely be limited to configurations that present a similar enough conformation to allow steric contact of important residues with target molecules. This limitation however, is probably mitigated to some extent by the inherent disordered structure of the loops, allowing a high degree of flexibility and hence, target seeking capability.

In order to investigate the functionally important structural features of CRPs and determine the species specificity of these proteins and their structural features, chimeric protein sequences were engineered. Due to the similar functional roles and high degree of sequence similarity, *Zm*MLP1 and *At*ESF1.3 were assessed to determine structural similarities at the sequence level. The amino acid sequences were aligned both with

and without cysteine anchoring (see 2.6.1) and annotated with the secondary structure data obtained in Figure 5.5 (a) and from the published work of Costa *et al* (Costa *et al.* , 2014) (Figure 5.8). This alignment was then used in conjunction with secondary structure prediction to identify structural similarities and differences, particularly the equivalent sequence loops that interspace the disulphide bonds (see 2.6.2).

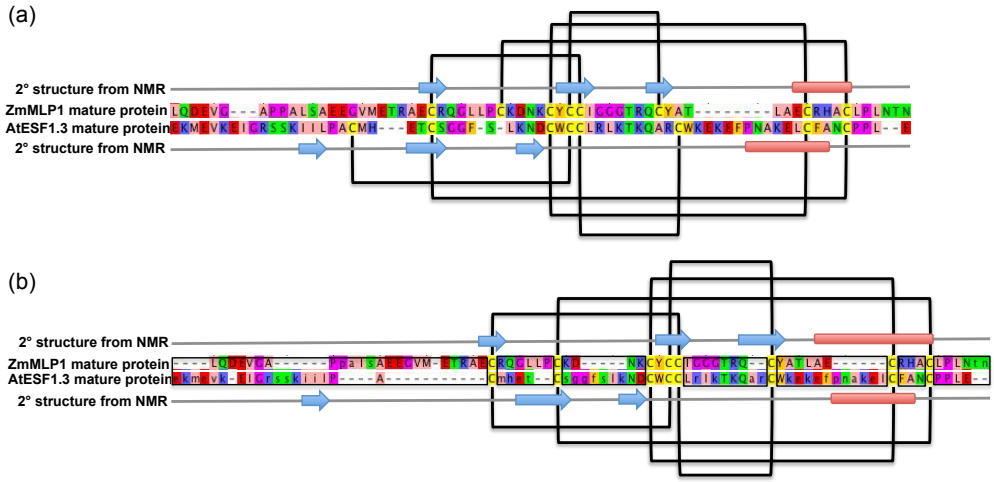

Figure 5.8: Identification of *Zm*MLP1 structural loops (a) Amino acid sequence alignment of *Zm*MLP1 and *At*ESF1.3 using the ClustalW algorithm. (b) Sequence alignment using the Clustal algorithm with cysteine anchoring performed by DIALIGN (Morgenstern *et al.* , 2006). Sequences annotated with NMR-derived secondary structure. β-sheets (blue arrows), α-helices (red cylinders) and disulphide bonds (black lines). Inter-cysteine loop sequences are indicated by black boxes. Amino acid residues coloured according to physiochemical properties.

Sequence chimeras were then made in a similar strategy to that used in the identification of the structural determinants of Insulin-Like Peptide 3 (Bathgate *et al.* , 2012). The *At*ESF1.3 sequence was used as a scaffold to support the substitution of the structural loops that had been identified in both proteins. In this way, constructs were engineered comprising the signal peptide of *At*ESF1.3 followed by a chimeric sequence in which a loop from *At*ESF1.3 has been replaced by the corresponding loop from *Zm*MLP1 (Figure 5.9 (c)). These constructs were then synthesised, cloned into the vector pMDC

ON2.1, transformed into *Agrobacterium* (see 2.3.7 and 2.1.10) and used to transform wild type *Arabidopsis* by floral dip (see 2.1.12).

The resulting T_0 transgenic seed was then selected for glufosinate resistance (see 2.1.13) and self pollinated to generate the T_1 generation. T_1 seed was then quantitatively selected by glufosinate resistance on media plates in order to estimate the number of transgene insertions by calculation of resistance frequency (see 2.1.15). Lines exhibiting a single insertion event were then grown to maturity and crossed with *Arabidopsis* lines expressing the pEBE-2:LhG4 trans-activator. This two-component system uses a transgenic activator line expressing a synthetic transcription factor, LhG4 under the control of the tissue specific promoter for the *AtEBE-2* gene (Costa *et al.* , 2014). The synthetic pOp promoter in the pMDC ON2.1 plasmid, is trans-activated by LhG4 when the plants are crossed, driving expression of the construct in the micropylar endosperm (Costa *et al.* , 2014; Moore *et al.* , 2006). In this way, tissue specific over-expression of the *At*ESF1.3-*Zm*MLP1 substitution constructs in ESR cells can be achieved.

Published work has shown that the over-expression of *At*ESF1.3 causes an abnormal elongation of the embryo suspensor in the early stages of embryo development (Costa *et al.* , 2014). In order to investigate if there are any functionally common features between *Zm*MLP1 and *At*ESF1.3, the substitution constructs were over-expressed in the micropylar endosperm and the embryos analysed to determine if the suspensor elongation phenotype could be induced. This was done by crossing the plants bearing the substitution constructs into LhG4 driver plants (pGRP-GUS:GFP) (Costa *et al.* , 2014) (see 2.1.6). Using a suspensor GFP marker, the seeds were then analysed by confocal microscopy to determine if there had been any effect on the length of the embryonic suspensor (see 2.5.5).

Upon completion of the 3D NMR structure, analysis of the models revealed some similarity between *Zm*MLP1 and its *Arabidopsis* homologue *At*ESF1.3 (Figure 5.9 (a and

b)). The overall structure in both cases, contains a Knottin motif formed from an alpha helix, three anti-parallel beta sheets and 6 disordered connecting loop regions held together by a disulphide bonded core scaffold. However, there are also some marked differences in the structure. The helix formed in *Zm*MLP1 is very short, consisting of only one turn and is positioned on the opposite face of the protein compared to *At*ESF1.3. This has the effect of forming something of a mirror image in the overall structure (Figure 5.9). Additionally, the loop regions are of varying lengths between the proteins and in some cases in different positions than predicted in Figure 5.8 due to some of the cysteines bing involved in beta sheets. This shifting of the loops in the verified 3D structure (Figure 5.9) changed the validity of some of the chimeric sequence designs which were originally engineered based on the predicted sequences alone given that the 3D structure was not yet available.

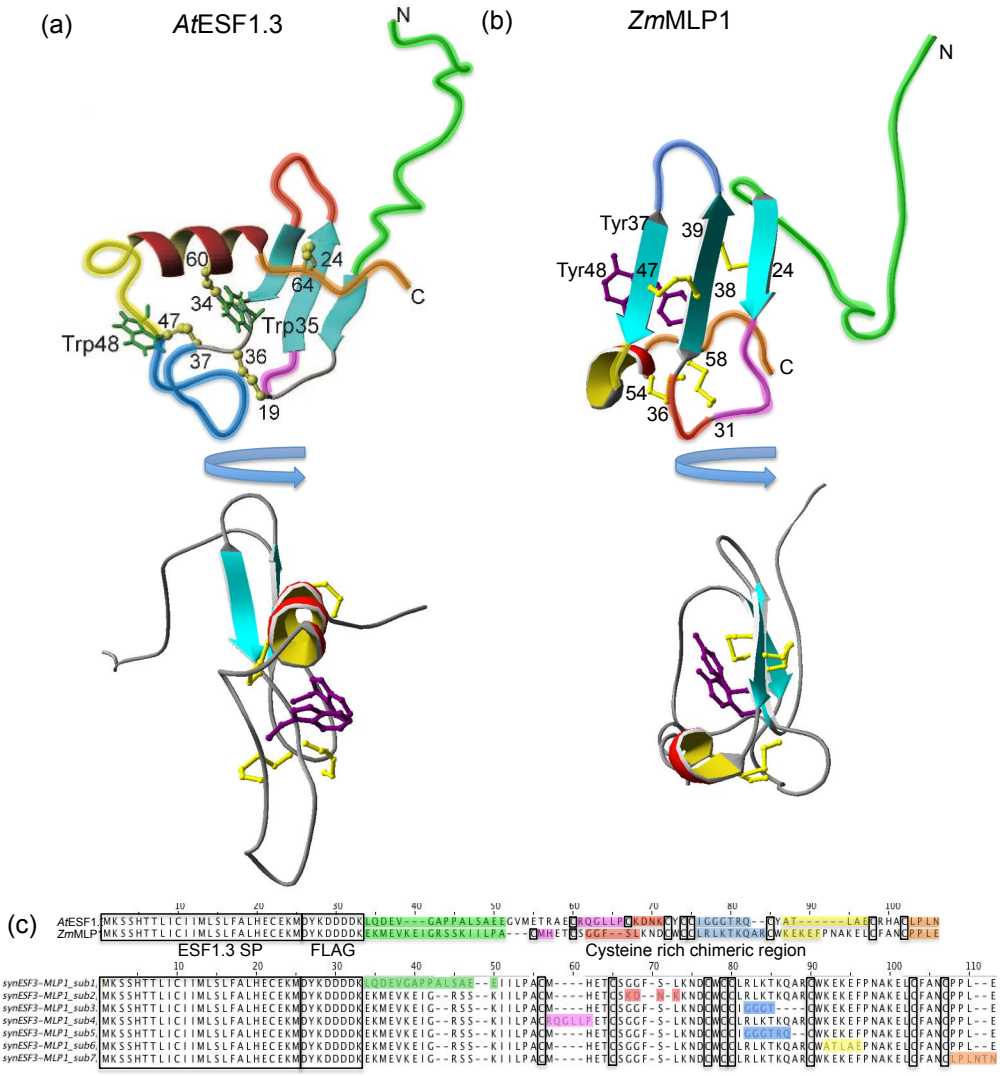

Figure 5.9: Generation of chimeric proteins. (a and b) NMR derived ribbon models of *At*ESF1.3 and Zm*MLP1* respectively in two orientations rotated about the y axis by 90º. Equivalent inter-cysteine loops as identified from the sequence are colour coded between the two structures, demonstrating the differences between the predicted and NMR solved loop positioning. (c) Sequence alignments of *At*ESF1.3 and Zm*MLP1* indicating the designated sequence loops together with alignments of chimeric *At*ESF1.3-Zm*MLP1* substitution constructs. All sequences use the *At*ESF1.3 signal peptide and sequence, separated by the FLAG recognition sequence as a scaffold with the substituted loops indicated using the same colour code as in (a) and (b). Green = loop 1, pink = loop 2, red = loop 3, blue = loop 4, yellow = loop 5, orange = loop 6.

The GFP marker in the chimeric lines allowed clear visualisation of the suspensor,

facilitating its measurement by image analysis (see 2.7.5). Similarly to the effect of over-expressing *At*ESF1.3, the over-expression of some substitution constructs caused an increase in the suspensor length compared to the wild type control (Figure 5.10). Upon analysis of all the lines, a significant increase in suspensor length was found in the plants expressing *At*ESF1.3, *At*ESF1.3$_{\text{-}53:54}$*Zm*MLP1$_{+58:63}$, *At*ESF1.3$_{\text{-}71:79}$*Zm*MLP1$_{+73:79}$ and *At*ESF1.3$_{\text{-}98:101}$*Zm*MLP1$_{+92:97}$ compared to the wild type (Figure 5.11 and Table 5.3).

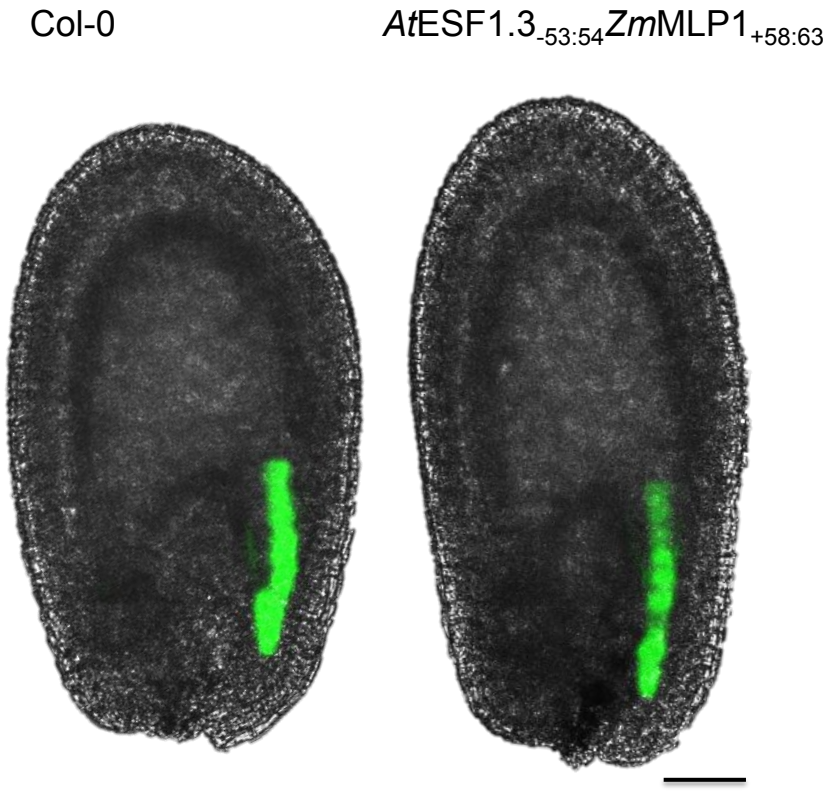

Figure 5.10: Phenotype observed in *Arabidopsis* suspensor cells expressing functional *At*ESF1.3-*Zm*MLP1 substitution constructs. Suspensor cells are visualised by GFP under confocal microscopy (green). (Left) Columbia-0, (right) *At*ESF1.3$_{\text{-}53:54}$*Zm*MLP1$_{+58:63}$. Scale bar = 50 µm

The chimeric line *At*ESF1.3$_{\text{-}34:46}$*Zm*MLP1$_{+34:48}$which has a substitution in the N-terminal lead sequence did not present an elongated suspensor phenotype (see Figure 5.11), in-

dicating that the protein chimera was not active. This suggests that the N-terminus plays a functional role in *At*ESF1.3 either through general physiochemical properties, or due to the presence of specific amino acids that could be involved in binding with an interacting partner. Loop 2 as substituted in line *At*ESF1.3$_{\text{-}59:63}$*Zm*MLP1$_{+65:68}$also appears to be functionally important to *At*ESF1.3, since the chimeric protein was not able to induce the phenotype (Figure 5.11). However, this substitution is not useful in determining cross-species compatibility, since the 3D structure revealed that the loops identified by sequence analysis do not align correctly due to the position of the disulphide bonds (Figure 5.9). Constructs *At*ESF1.3$_{\text{-}71:79}$*Zm*MLP1$_{+73:77}$ and *At*ESF1.3$_{\text{-}71:79}$*Zm*MLP1$_{+73:79}$yielded more interesting results given that the substitution is in the same loop but to different extents. Interestingly *At*ESF1.3$_{\text{-}71:79}$*Zm*MLP1$_{+73:77}$was unable to induce the phenotype while *At*ESF1.3$_{\text{-}71:79}$*Zm*MLP1$_{+73:79}$showed chimera activity (Figure 5.11). The only difference between the substitutions was the extension of the donor *Zm*MLP1 loop to include *Zm*MLP1$_{\text{R}_{78}}$ and *Zm*MLP1$_{\text{Q}_{79}}$ in place of *At*ESF1.3$_{\text{K}_{78}}$ and *At*ESF1.3$_{\text{Q}_{79}}$. Although the loops are not aligned according to the 3D structure (Figure 5.9), the amino acids are extremely similar in their side chain properties - glutamine is obviously an identical swap, while lysine and arginine are both positively charged residues, thus, maintaining the polar nature of this highly exposed sequence loop. Moreover, despite the opposite orientation of the protein structure with regards to the beta sheet, this loop remains in the same relative position to the hydrophobic array previously mentioned.

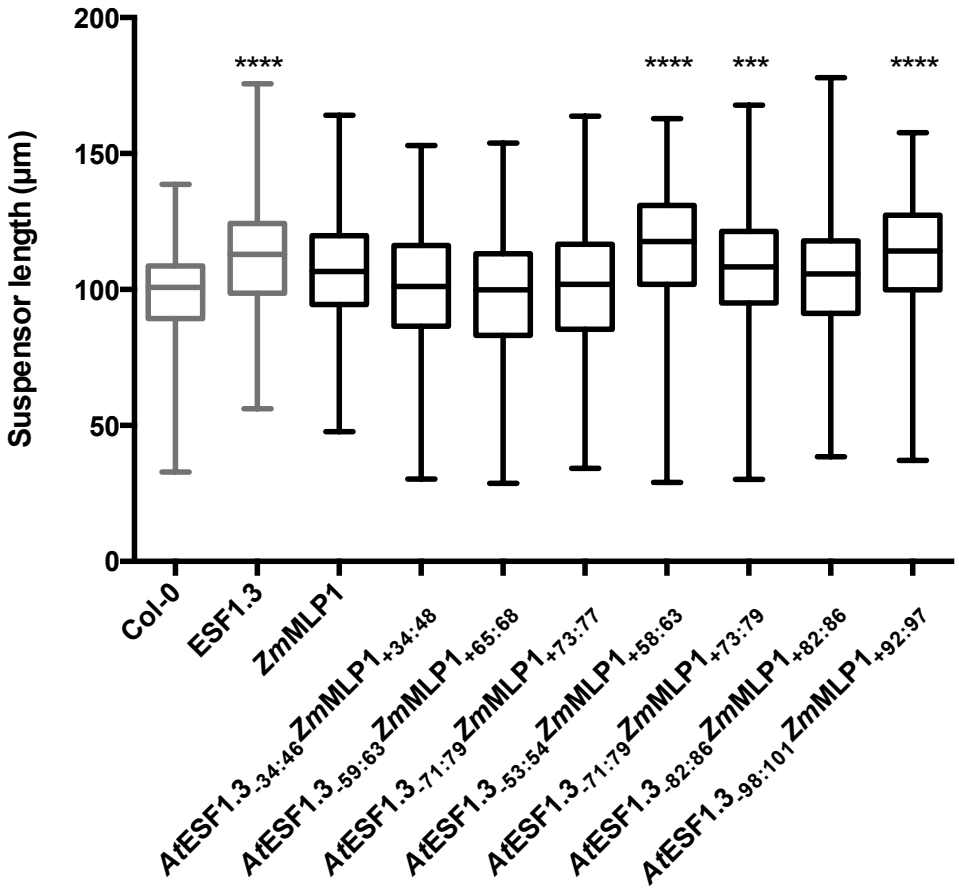

Figure 5.11: Effect of expression of *At*ESF1.3-*Zm*MLP1 substitution constructs in the micropylar endosperm on the length of the embryo suspensor. Box and whisker plots show the mean length of the suspensor in lines expressing the different substitution constructs (black) against the wild type and *At*ESF1.3 over-expression controls (grey). *** and **** = p-values of <0.001 and <0.0001 respectively.

The line *At*ESF1.3$_{-53:54}$*Zm*MLP1$_{+58:63}$ showed an increased length in suspensor (Figure 5.11) indicating that the substituted protein was active. This again indicates that either the amino acids removed from *At*ESF1.3 were not of functional importance, or that the donor residues are similar enough to achieve the same biochemical properties, although this does not seem likely from the side chain chemistry. Additionally, this substitution has increased the length of the loop, suggesting that the protein is not

Table 5.3: One-way ANOVA multiple comparison test (Dunnett's) of suspensor lengths across different substitution lines and control crosses.

Line	**Mean suspensor length (µm)**	**n**	**Mean Difference against Col-0**	**95 % CI of MD**	**Significance**
Col-0	98.64	133	-	-	-
*At*ESF1.3	112.8	174	14.17	20.58 to 7.764	****
*Zm*MLP1	106.4	260	7.749	13.68 to 1.819	**
*At*ESF1.3$_{-34:46}$ *Zm*MLP1$_{+34:48}$	99.57	136	0.938	7.722 to -5.846	ns
*At*ESF1.3$_{-59:63}$ *Zm*MLP1$_{+65:68}$	96.22	161	-2.421	4.097 to -8.940	ns
*At*ESF1.3$_{-71:79}$ *Zm*MLP1$_{+73:77}$	99.51	188	0.873	7.176 to -5.431	ns
*At*ESF1.3$_{-53:54}$ *Zm*MLP1$_{+58:63}$	115.6	213	16.93	23.08 to 10.79	****
*At*ESF1.3$_{-71:79}$ *Zm*MLP1$_{+73:79}$	106.9	340	8.269	13.96 to 2.579	***
*At*ESF1.3$_{-82:86}$ *Zm*MLP1$_{+82:86}$	104.6	193	5.953	12.22 to -0.3168	ns
*At*ESF1.3$_{-98:101}$ *Zm*MLP1$_{+92:97}$	113.1	428	14.49	20.01 to 8.967	****

sterically sensitive in this particular region.

*At*ESF1.3$_{\text{-}82:86}$*Zm*MLP1$_{+82:86}$failed to induce the extended suspensor phenotype, suggesting that the changes made to *At*ESF1.3 were functionally prohibitive. This could be due to the removal of a charge-balanced sequence loop and the introduction of hydrophobic residues which could have an effect on molecular interactions. The final construct tested, *At*ESF1.3$_{\text{-}98:101}$*Zm*MLP1$_{+92:97}$, swapped the C-terminal tail between the proteins and resulted in an active chimera (Figure 5.11). This suggests that either the tail is not functionally important, or that the residue properties are compatible. At the sequence level, the tails are initially chemically similar, however, the substitution introduces a highly polar terminus from *Zm*MLP1, though this appears to have no deleterious functional effect.

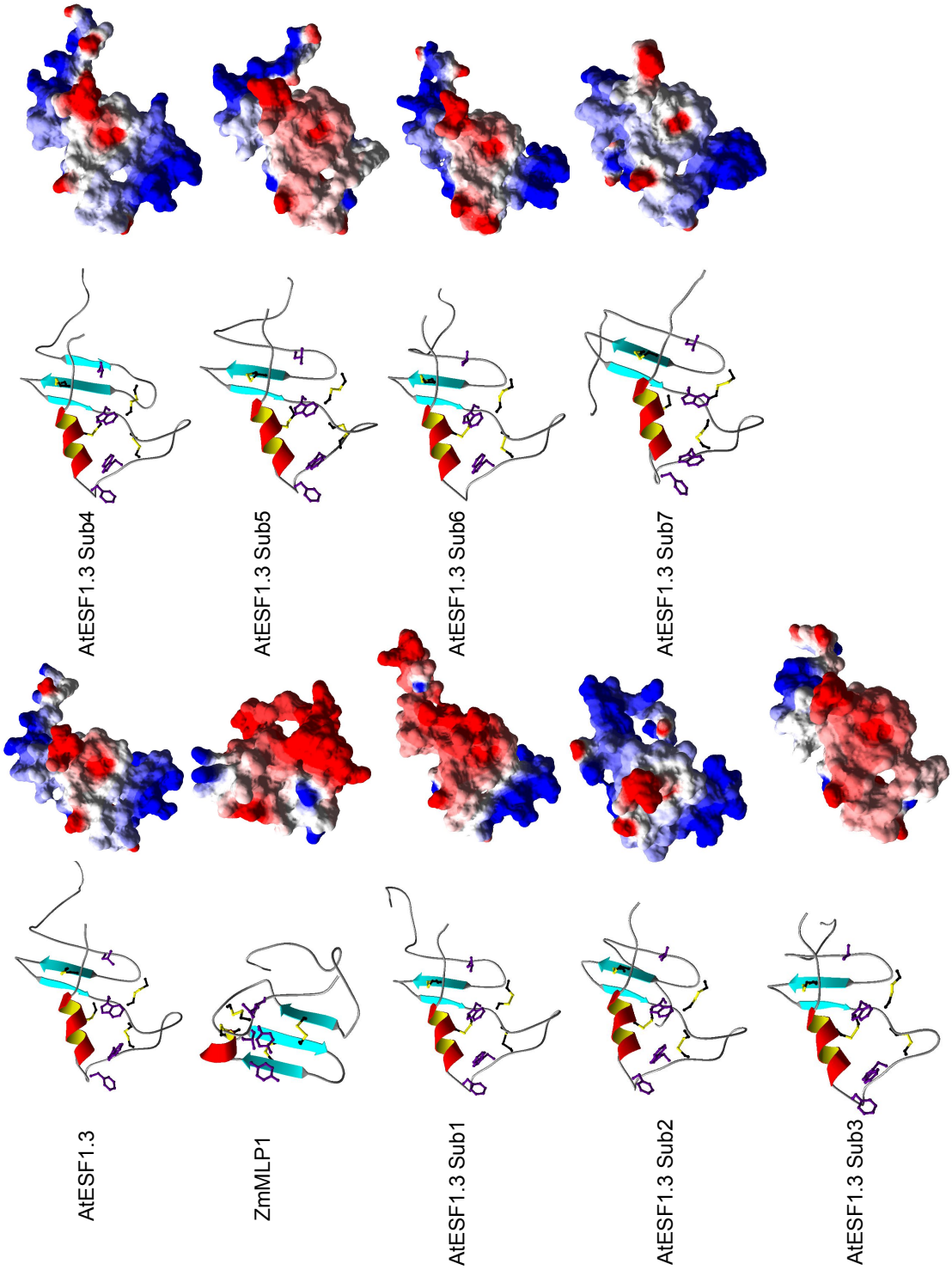

Figure 5.12: Homology based (left) ribbon and (right) surface models of the chimeric *At*ESF1.3-Zm*MLP1* substitution constructs showing altered secondary structure and electrostatic potential distribution respectively. (red) Negatively charged , (blue) positively charged , and (grey) hydrophobic residues.

In further collaboration with Professor Ohki, homology based modelling was carried out on the chimeric proteins, using the *Zm*MLP1 and *At*ESF1.3 structures as templates. In this way, 3D structural and surface models were generated for each of the chimeras, demonstrating the structural and chemical consequences for the amino acid substitutions (Figure 5.12). An obvious chemical change in the protein surface was observed due to the differing side-chain chemistry of the substituted sequence loops. This was manifested as a significant alteration in both the charge distribution and hydrophobicity of patches on the molecular surface between the different chimeras. Due to the avoidance of substituting sequences directly involved in secondary structure, the core scaffold of the proteins have remained relatively unchanged and the secondary structure largely undisturbed.

An interesting observation in the modelling experiments was the presence of two additional hydrophobic residues, Leu_{16} and Phe_{53} in the conserved aromatic array in *At*ESF1.3. Phe_{53} is missing in *AtESF1.3-82:86ZmMLP1+82:86*, which has caused a strong shift from hydrophobic to negative charge in the electrostatic surface potential in a distinct region, potentially contributing to the lack of activity of this construct (Figure 5.11). This presents evidence that the aromatic array may play a crucial role in the activity of these proteins, but as seen by the lack of activity in *At*ESF1.3$_{\text{-}34:46}$*Zm*MLP1$_{+34:48}$, *At*ESF1.3$_{\text{-}59:63}$*Zm*MLP1$_{+65:68}$ and *At*ESF1.3$_{\text{-}71:79}$*Zm*MLP1$_{+73:77}$, which all contain the array, other structural features must be playing a part. The loop substitutions in these constructs have had a significant effect on the electrostatic surface potential compared to *At*ESF1.3, which could have an impact on hydrophobic associations with interacting proteins. As previously mentioned, *At*ESF1.3$_{\text{-}71:79}$*Zm*MLP1$_{+73:77}$ is inactive, while *At*ESF1.3$_{\text{-}71:79}$*Zm*MLP1$_{+73:79}$ is active while the only difference between the two constructs is the inclusion of two amino acids, Arg and Gln, in loop 4. These two amino acids are also present in the wild type *At*ESF1.3 loop and all the active

chimeras, presenting evidence that these amino acids (or their positive and polar properties, respectively) may be functionally critical. This property change is made evident by the change in surface properties illustrated in the models in Figure 5.12.

5.6 Discussion

*Zm*MLP1 is a typical example of a small CRP, bearing many structural similarities to the published *At*EPFL9 (Ohki *et al.* , 2011), *Br*SP11 (Mishima *et al.* , 2003) and *At*ESF1.3 (Costa *et al.* , 2014) proteins. The generation of a reliable structure for *Zm*MLP1 has added further weight to the mounting evidence for a conserved structural format in the small CRP protein class. This conformational scaffold presents an evolutionarily valuable tool for the development of stable, small proteins with a potentially enormous functional variability. It seems to be clear that the functionality of the proteins is conferred by the disulphide rich scaffold core, supporting a variable and probably function-specific superstructure. Furthermore, the conservation of signal peptides in this protein class could be instrumental in upholding the role specificity of its members, based on the direction of proteins to spatially relevant tissues or cell compartments. The small size and observed secretion potential of CRPs makes them ideal candidates for signalling molecules involved in cell-cell communication.

The transferrable nature of functional loops between the maize and *Arabidopsis* homologues gives additional insight into the operational mechanisms used in the CRP protein class. The disulphide bonds have been shown previously to be crucial in CRP structural efficacy, the correct pairing of which is essential for proper protein conformation (Costa *et al.* , 2014; Ohki *et al.* , 2011). The NMR structure of *Zm*MLP1 and the modelling data from the chimeric proteins indicate that secondary structure is probably important for the correct spatial arrangement of the functional residues. However, given that the

secondary structure was largely undisturbed in the *At*ESF1.3 chimeras, but function was lost in some cases, it is clear that the correct secondary structure is not sufficient. In conjunction with the *in planta* analyses, the modelling data suggested that the observed array of hydrophobic residues was functionally important to *At*ESF1.3, similar in arrangement to the aromatic array seen in *Zm*MLP1. This suggests that the conserved aromatics in *Zm*MLP1 and indeed *Zm*MEG1, may be structurally and functionally homologous to those in *At*ESF1.3.

Given the observations made here, is is becoming clearer that the linked-loop structural configuration of the CRPs is extremely adaptable. The amino acid composition of the loop region is likely to be the primary determinant of the functional evolution of this protein class. The loop substitution experiments presented in this chapter give evidence to the idea that the core scaffold of the CRP is extremely important in the functional stability of the protein, able to maintain activity in chimeric proteins with altered loop composition. Additionally, the composition of the loops confers the surface properties of the protein, the alteration of which can cause loss of function. It would be interesting to carry out point substitution experiments on both *Zm*MLP1 and *At*ESF1.3 to determine more accurately the functionally critical residues and whether they are specifically required or could be replaced with residues of a similar physiochemical status. This could be conducted by a larger scale experiment in which the loops of interest could be either systematically probed by the substitution or deletion of individual amino acids, or by the generation of a library of peptides with randomised loop sequences. Furthermore, the additional amino acids substituted in could be chosen for particular side chain properties to investigate the importance of protein surface properties in activity. It would also be interesting to reverse the substitution strategy, introducing *Arabidopsis* loops into the maize protein and testing the new chimeras is maize and *Arabidopsis* to see if an effect could be induced. However, the generation of transgenic maize lines is

technically challenging, laborious and expensive using the methods that have been used in this work.

A further observation made in the structural assessment of *Zm*MLP1 was the degree of flexibility and disorder in the terminal and loop regions of the protein. This could be indicative of a conformational change in the protein upon binding with an interacting partner. Future work on this protein is to include binding partner identification through proteomics experiments and it would be interesting to solve the structure of any complexes identified to further investigate the relevance of the mobility of certain parts of the sequence. A related observation was the predicted phosphorylation sites on *Zm*MLP1, it would be interesting to investigate further the validity of these predictions and identify a possible kinase associated with this phosphorylation. A significant element in this hypothesis is the link back to the differential gene analysis conducted in chapter 4, which identified an enrichment in kinases in the up-regulated gene cluster associated with *Zm*MLP1 down-regulation. This opposing relationship in expression could be due to a negative feedback loop as part of a regulatory system for the associated kinase. This theory however is merely speculative at this stage as phosphorylation site prediction can be unreliable as it is based only on sequence effects.

6 Conclusions and Further Work

Small CRPs form an interesting, diverse and rapidly expanding protein class with a wide range of functions, interest in which has formed a progressively important field of research. Plants in particular have yielded a highly abundant compendium of CRPs with roles in defence, signalling and perception, development and reproduction but to name a few (see Table 1.1). As a globally important food crop and economic commodity, maize is becoming an increasingly important model crop plant to study in the fields of biotechnology and crop science. For obvious reasons, the research niche of reproductive development is of vital importance to the agricultural and biotechnological industries, since these processes are directly related to the yield potentials available in modern farming.

6.1 *Zm*MLP1 is an evolutionarily conserved ESR factor

One of the major aims of this project was to characterise the expression of a novel small CRP, identified in the maize genome by studying its promoter activity, spatio-temporal gene expression profile, protein localisation and biochemical properties. *Zm*MLP1 was found to exhibit a tightly controlled expression pattern, temporally and spatially regulated in the ESR of the developing seed. Interestingly, expression was detected in the unfertilised ovule, restricted to the central cell as demonstrated by both GUS reporter and GFP fusion experiments. After fertilisation, *Zm*MLP1 is expressed at high levels throughout lifespan of the ESR in the developing seed. This conclusion is supported by temporal and spatial gene expression analysis, promoter analysis using a GUS reporter and protein localisation by GFP fusions and immunological detection methods. Furthermore, evidence was gathered that demonstrated the function of the cleaved signal peptide in the secretion of *Zm*MLP1 into the apoplast.

The observed level of regulatory control exhibited by *ZmMlp1* with regards to its tem-

poral and spatial restriction was strikingly different from the expression pattern displayed by the other members of the *ZmMEG1* gene family such as *ZmMeg1* (Gutierrez-Marcos *et al.* , 2004), probably due to the differences observed in the promoter regions. *ZmMlp1* presents an expression pattern that partially aligns with other ESR genes such as *ZmEsr1-3*, *ZmESR-6*, *ZmAE1* and *ZmAE3* (Opsahl-Ferstad *et al.* , 1997; Bonello *et al.* , 2000, 2002; Balandín *et al.* , 2005; Magnard *et al.* , 2000), with the additional feature of central cell expression. Similarly, *ZmMlp1* partially shares the transcriptional pattern of *ZmEBE* which is expressed in the central cell, ESR and BETL (Magnard *et al.* , 2003), though *ZmMlp1* shows no BETL expression. Thus, *ZmMlp1* has been shown to exhibit a novel expression pattern in maize, localising exclusively to the central cell and ESR.

Interestingly, a recent publication has identified the *ZmMlp1* gene (referred to as Meg13) in an analysis of the evolutionary expansion of the Meg1 family (Xiong *et al.* , 2014). Through phylogenetic sequence analysis, Xiong *et al* have determined that *ZmMlp1* may have diverged from the Meg1 family before the divergence of the sorghum and maize species, adding weight to the thought that *ZmMlp1* is an ancestral gene which gave rise to the Meg1 family in maize, as hypothesised from our own sequence analysis.

The analysis by Xiong *et al* suggested that the secondary structure of the Meg1 family proteins have adapted with their expansion, presumably into discrete functional roles. However, these observations were based solely upon evidence obtained from *in silico* structural prediction, a process that can be unreliable. In fact, the secondary structure of *Zm*MLP1 (Meg13) predicted by Xiong *et al*, is very dissimilar to the NMR structure obtained in my work. Furthermore, the expression analysis of the Meg1 family conducted by Xiong *et al* failed to identify any expression of *ZmMlp1* (Meg13) in the seed due to sampling being carried out in different endosperm tissue types and at developmental time points that were too late for *ZmMlp1* expression.

In order to further understand the expression pattern of *ZmMlp1* and its homologues, a more detailed promoter analysis needs to be conducted. The level of constraint in the spatial and temporal displayed by *ZmMlp1* could represent a valuable tool in the biotechnology industry for the specific expression of useful transgenes in the developing seed. A deeper analysis at the sequence level of ESR specific genes, combined with reporter genes driven by suspicious fragments could potentially identify a unique promoter element responsible for the observed expression pattern. In this way, synthetic promoters could be designed, driving expression in the ESR of any useful genes. It would also be interesting to investigate further the sub-cellular localisation of the mature protein. While evidence has been presented here in chapter 3 that demonstrates the secretion of the protein into the apoplast, experimental evidence for this process in the native maize ESR cells is still lacking. One approach to verify this would be immuno-gold labelling electron microscopy, which could obtain a highly detailed characterisation of *Zm*MLP1 with regards to its location and associations within the cell structure.

6.2 *Zm*MLP1: Structure-function relationships

A second aim of this research was to discover the key structural features of *Zm*MLP1 through structural determination and use the information gained to conduct structure-function based experiments that investigate the importance of the CRP protein class. The structure of *Zm*MLP1 was determined by solution state NMR spectroscopy, conducted on a recombinant version of *Zm*MLP1, generated in a characterised plant cell expression culture to ensure correct protein folding. The revealed structure consisted of a disulphide-bonded core scaffold, supporting an ordered sequence of β-sheets and a small α-helix, which in turn, are interspaced by disordered loops that project outward forming the solvent accessible surface of the protein. This structure was observed to be

broadly similar to that of the *Zm*MLP1 homologue, *At*ESF1.3 in *Arabidopsis* (Costa *et al.* , 2014), and bears similar features to other published CRPs such as *At*EPFL9 (Ohki *et al.* , 2011) and *Br*SP11 (Mishima *et al.* , 2003).

In order to assess the functional importance of specific structural features in *Zm*MLP1, chimeric proteins engineered from the loops of *Zm*MLP1 and the scaffold of *At*ESF1.3 were assessed for functionality in transgenic *Arabidopsis* lines. In combination with modelling data, several structural features were identified which seem to bear some functional importance, including specific disordered loops and the presence of a conserved array of aromatic amino acid side chains which alter the physiochemical properties of the molecular surface. In order to make the conclusions drawn from these observations more robust, the analysis should be repeated in homozygous transgenic lines, which will be free from the mitigating effects of segregation. Furthermore, it would be interesting to repeat the experiment in reverse, using the *Zm*MLP1 protein as the scaffold, substituting loops in from *At*ESF1.3 to determine if the core is interchangeable, as long as the correct loop regions are present. Additionally, point mutation work would offer greater precision in the determination of key functional residues in both the structural integrity and functionality.

Linking to the key functional features of *Zm*MLP1, the generation of a functional GFP fusion offers the ability to conduct protein-protein interaction studies, through a proteomics approach using co-immunoprecipitation. In this way, associated proteins could be identified, and mapped to the corresponding amino acid sequences to determine the interacting surfaces of *Zm*MLP1. Identification of a receptor for example, could then lead to the discovery of a novel pathway, forming a true link between important structural features and the phenotypic function. This hypothetical pathway could be postulated further by analysis of the interacting genes identified in the RNAi based transcriptomic experiments carried out. On a broader note, it would also be interesting

to use circular dichroism to assess the structure of the recombinant *Zm*MLP1 used in the NMR experiments. This would lead to a greater understanding of the spectral output yielded by small proteins with secondary structure, but a high degree of interspaced disorder.

6.3 *Zm*MLP1 - A maternally required endosperm factor

This work also aimed to investigate the function of *ZmMlp1* through the phenotypic analysis of transcriptional knock-down transgenics, using an RNAi based approach. The functional characterisation of *Zm*MLP1 has offered the first robust role for an ESR peptide in maize. The reduction in expression of *ZmMlp1* demonstrated a novel approach to the functional characterisation of the ESR, demonstrating a key role in the normal development of the embryo. Under the condition of reduced *Zm*MLP1 concentrations in the ESR, the embryo suspensor failed to elongate to normal levels, together with morphological abnormalities observed in the embryo proper. This led to the conclusion that *Zm*MLP1 and therefore the ESR, must have a role in normal embryo pattern formation. This conclusion aligns well with the published data in *Arabidopsis*, demonstrating that the *Zm*MLP1 homologue, *At*ESF1.3, has a crucial role in elongation of the suspensor cells and normal patterning of the embryo proper (Costa *et al.* , 2014). As previously discussed, *At*ESF1.3 acts synergistically with a receptor-like kinase in the YODA kinase pathway, an observation which may give insight into the potential presence of a similar pathway in maize.

Interestingly, the genetic complementation experiments carried out in this study not only confirmed the functionality of the *Zm*MLP1-GFP fusion protein, but demonstrated the importance of the expression in the central cell implicating a role for *Zm*MLP1 before fertilisation or in the stages immediately after. This was shown by the inability

of the synthetic version of *Zm*MLP1 to fully recover the phenotypes exhibited by RNAi based knockdown of *Zm*MLP1 when delivered late, from the pollen. This information could be valuable for the generation of useful transgenics for the further study of ESR function as it is obviously important to consider the parental origin of the transgene.

A further observation from the RNAi studies, was the overall negative effect on seed size, weight and grain filling, linking *ZmMlp1* to crop yield. This could be evidence of cross-talk between the developing embryo and the surrounding endosperm in both directions. The developmentally retarded/altered embryo may place different demands on the endosperm for nutrient supply, potentially attempting to prevent the allocation of resources to an embryo which may not survive. Alternatively, the alteration in *Zm*MLP1 levels may have had an indirect effect on seed size due to the build up or lack of sugar signalling molecules caused by a defective ESR. Furthermore, we cannot rule out the possibility of off target effects of the RNAi approach, though every effort was made to reduce this possibility.

Recent advances in gene editing technologies such as TALENs (Boch *et al.* , 2009) and CRISPR/Cas (Mussolino & Cathomen, 2013) potentially offer an exciting new approach for functional gene studies through more targeted gene interuption. Application of these techniques has already been successfully applied to a variety of plant species (Puchta & Fauser, 2013) with recent work demonstrating specific gene editing on a multiple target genes in maize (Liang *et al.* , 2014). This methodology has the potential to revolutionise the way in which loss of function experiments are carried out, in addition to presenting a new way of making targeted alterations to a protein sequence for structure-function analyses.

In order to further characterise the function of *Zm*MLP1, an over-expression line would be useful. It would be interesting and potentially commercially valuable to determine if the over-expression of *Zm*MLP1 could cause the opposite effects, leading to improved

crop yield. As previously discussed, the identification of binding parters and factors involved in a potential pathway would be highly useful to determine if this gene is a useful biotechnological asset.

In summary, this work has contributed to the understanding of plant reproductive development, the role of CRPs in the maize seed and the structural determinants of CRP functionality. The characterisation of *Zm*MLP1 has revealed a novel plant CRP with a unique expression pattern in maize. A functional role in embryonic development has been assigned to *Zm*MLP1 and its structure has been investigated, leading to a deeper understanding of CRPs, their roles in plants and their mechanisms of action. This study has shown that *Zm*MLP1 is a maternal factor required for embryo development in maize and bears functional resemblance to endosperm specific CRPs in other species. Additional work in this area will likely contribute to protein engineering and the fundamentals of plant reproductive development. Finally, *Zm*MLP1 has been shown to be a regulator of maternal-to-offsping nutrient allocation, a characteristic potentially useful in biotechnological crop improvement.

7 Appendix

Contents

7.1 Synthetic gene constructs

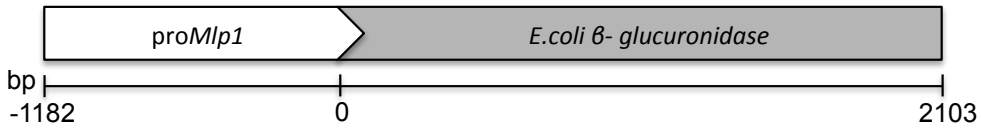

Figure 7.1: Schematic representation of the MLP1-GUS fusion construct used to drive GUS expression in transgenic maize.

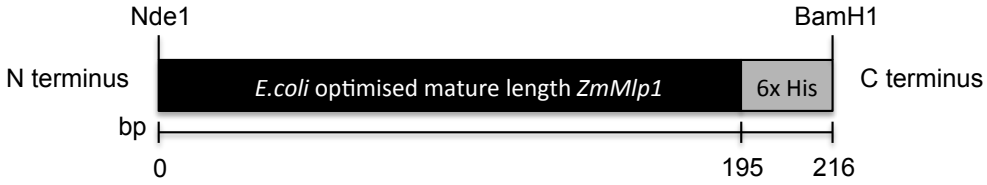

Figure 7.2: Schematic representation of the synthetic construct used to express His-tagged MLP1 in *E.coli.*

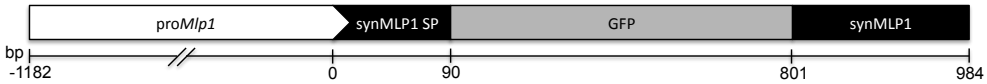

Figure 7.3: Schematic representation of the synthetic construct used to express His-tagged MLP1 in *E.coli.*

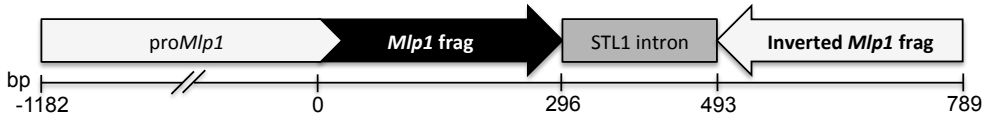

Figure 7.4: Schematic representation of the construct used to drive expression of a 296 bp ZmMlp1 fragment and an inverted repeat of the same fragment separated by the 197 bp potato *STL1* intron as a spacer.

Figure 7.5: Schematic representation of the synthetic construct used to express *Zm*MLP1 in tobacco. The full-length synthetic *ZmMlp1* gene was His-tagged and the PreScission protease recognition site included (PS). The sequence was flanked with AttB1/2 sites to facilitate Gateway cloning.

7.2 Vector Maps

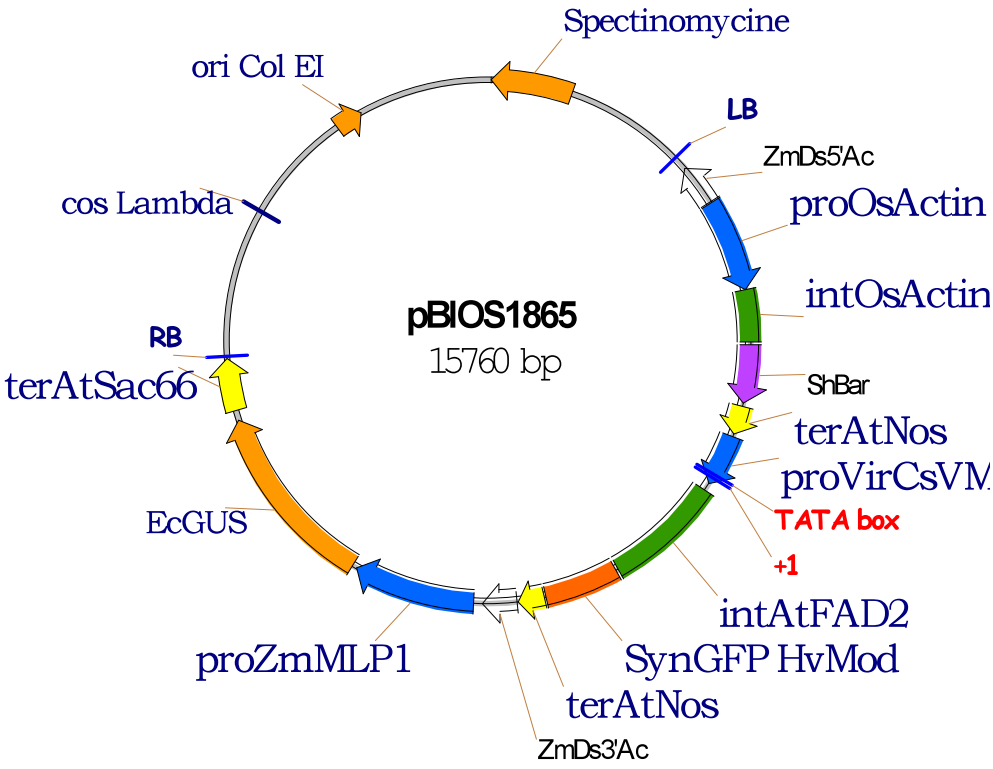

Figure 7.6: Vector map of pBIOS 1865 used to express pro*Mlp1*:GUS construct in transgenic maize plants.

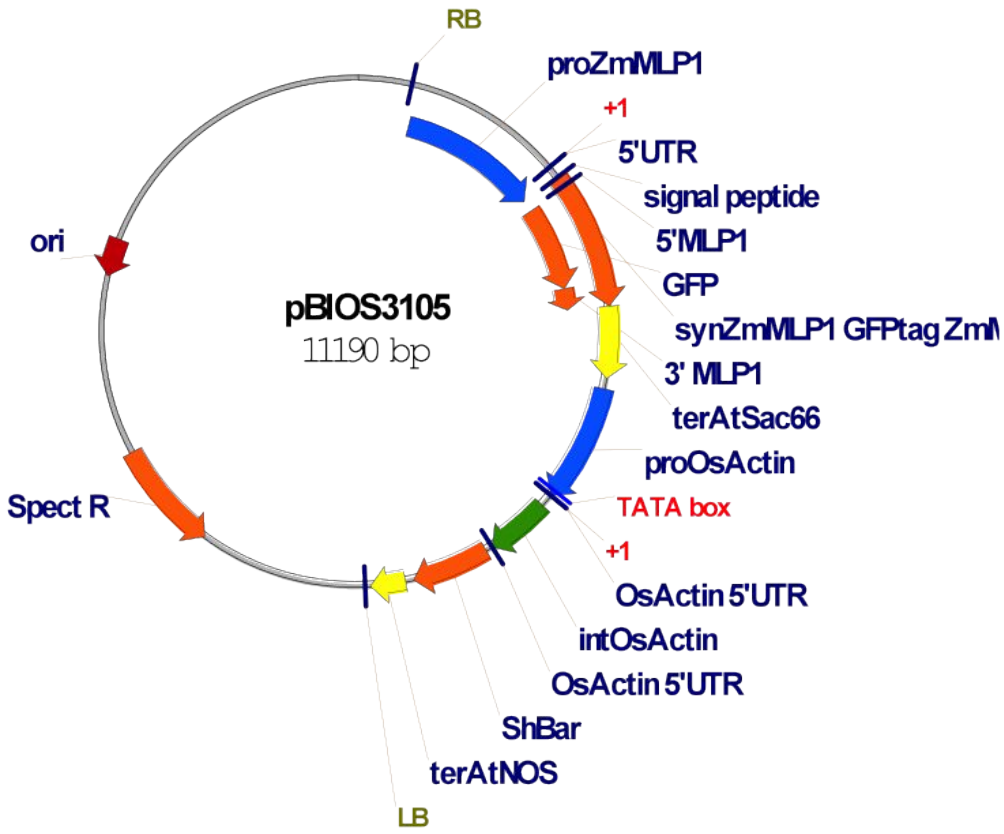

Figure 7.7: Vector map of pBIOS 3105 used to express MLP1:GFP construct, driven by pro*Mlp1* in transgenic maize plants.

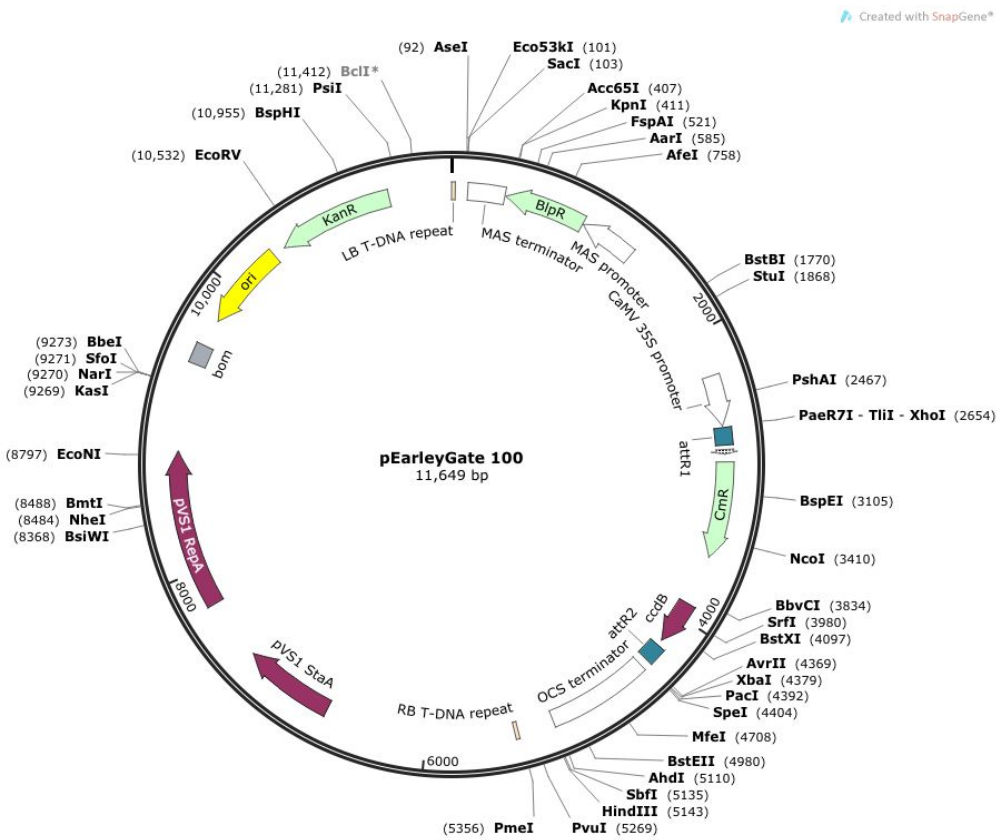

Figure 7.8: Vector map of pEARLEYGATE 100 used to express MLP1:GFP construct, driven by proCaMV 35S in *N.Benthamiana*.

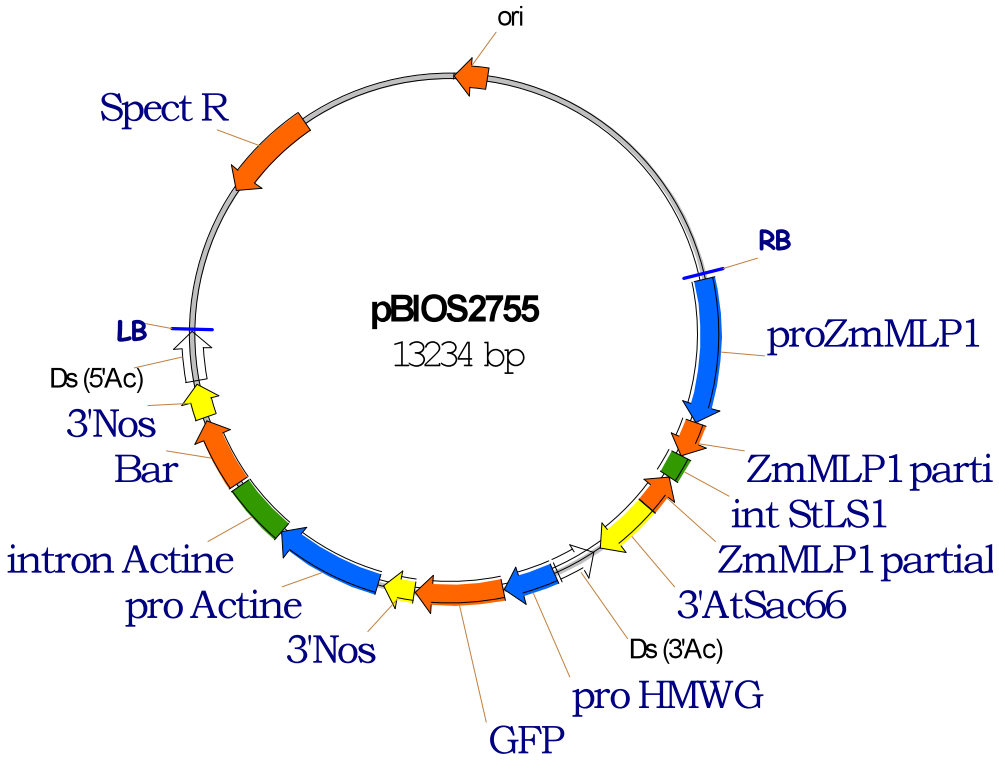

Figure 7.9: Vector map of pBIOS 2755 used to express MLP1 RNAi construct, driven by pro*Mlp1* in maize.

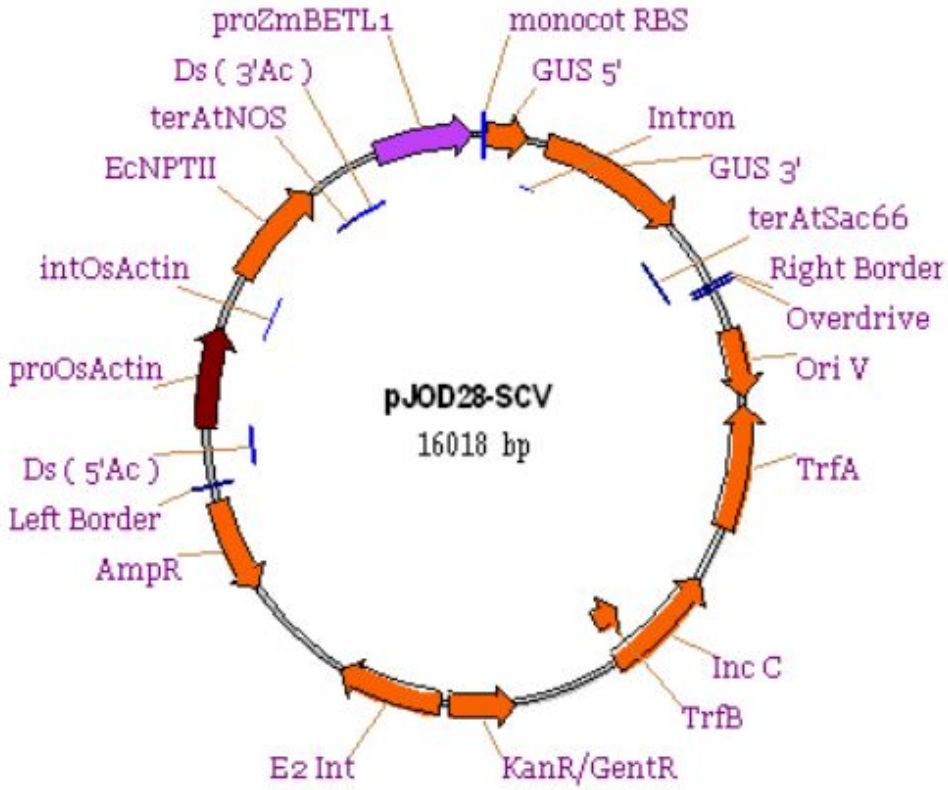

Figure 7.10: Vector map of pJOD28-SCV used to express pro*ZmBETL*1:GUS in wheat.

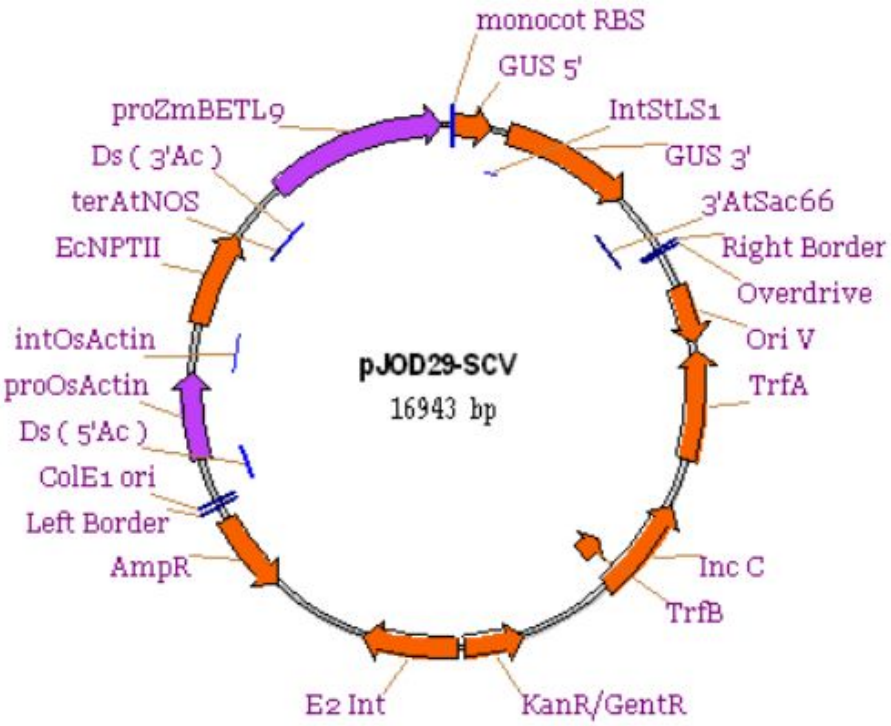

Figure 7.11: Vector map of pJOD29-SCV used to express pro*ZmBETL9*:GUS in wheat.

7.3 Sequence Alignments

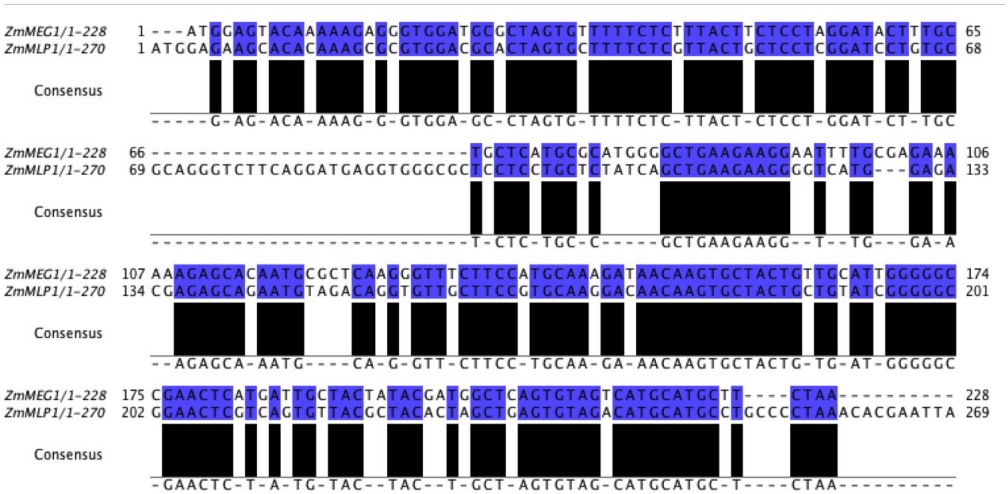

Figure 7.12: MLP1 vs MEG1 DNA sequence alignment

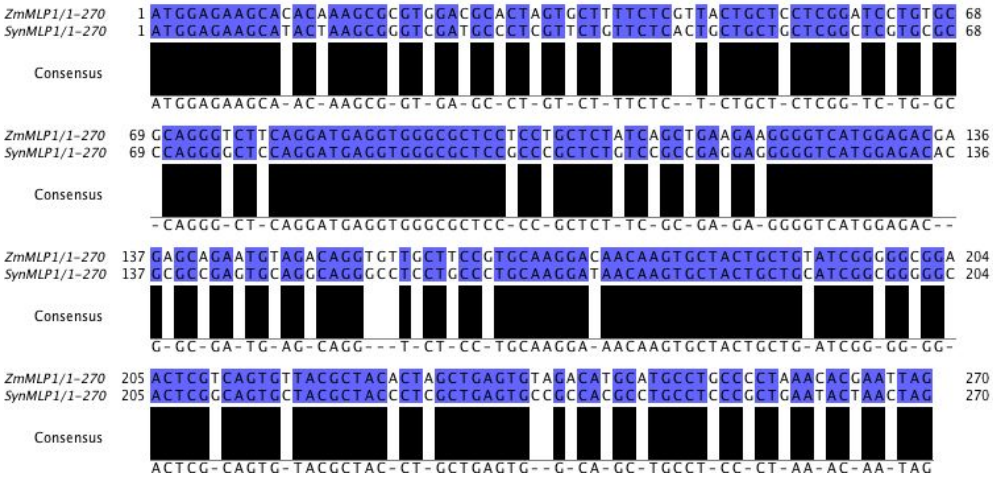

Figure 7.13: MLP1 vs synthetic MLP1 DNA sequence alignmentnMLP1

7.4 PCR primer sequences

Table 7.1: Genotyping PCR primer sequences

Gene	Primer Name	Sequence (5' - 3')	Size (bp)
Bar	BAR fwd	CCATCGTCAACCACTACATCGAG	322
	BAR rev	TTGGGCAGCCCGATGACAGCGAC	
Yfp	qYFP fwd	GTGAACCGCATCGAGCTGAAGG	91
	qYFP rev	CGTTGTGGCTGTTGTAGTTGTACTCC	
syn Meg1-PH	synMEG1-PH fwd	GATGATGTCAATGTTTCTGCTCCAG	210
syn Mlp1-PH	synMLP1-PH fwd	CCTCCTGCTCTATCAGCTGAAGAAG	207
	synMLP1/MEG1-PH rev	GTGATGGTGATGGTGCAATTCTACT	
T7e	T7e fwd	GTGAACCGCATCGAGCTGAAGG	91
	qYFP rev	CGTTGTGGCTGTTGTAGTTGTACTCC	

Table 7.2: qRT-PCR primer sequences. Introns marked with *

Gene	Primer Name	Sequence (5' - 3')	Size (bp)
Mlp1	MLP1 exon 1 fwd	CTCGTTACTGCTCCTCGGATCC	416 (gDNA)-109 (cDNA)
	MLP1 exon 2 rev	CTACATTCTGCTCTCGTCTCCATGA	
Mlp1	MLP1 qRT fwd	GCAGGGTCTTCAG*GATGAGGTG	85
	MLP1 qRT rev	CTGTCTACATTCTGCTCTCGTCTCCA	
Mrp1	MRP1 qRT fwd	GACTACAGATGAGCACAG*GAATTTC	120
	MRP1 qRT rev	GCATGGCTAGAGATCTGCA	
Esr6	ESR6 qRT fwd	GCCATAACCATGCCGTCCT	90
	ESR6 qRT rev	TGCAGACGCATCCATTC*CGA	
Al9	AL9 qRT fwd	CTATGTTTGCCATAGGCTCTCATGC	283
	AL9 qRT rev	GCTGGAACCTTGTAGC*TTCCG	
Gapdh	GAPDH qRT fwd	GTCACAGATGGTAGCAGGAAGGGAAG	120
	GAPDH qRT rev	GTGTATGCCGAGAATAAATGTGGATG	

Table 7.3: Gateway cloning primer sequences. attB sequences in lower case

Gene	**Primer Name**	**Sequence (5' - 3')**	**Size (bp)**
Mlp1-Gfp	MLP1-GFP attB1 fwd	ggggacaagtttgtacaaaaaagca-ggcttcATGGAGAAGCATA-CTAAGCGG	984
	MLP1-GFP attB2 rev	ggggaccactttgtacaagaaagctg-ggtcCTAGTTAGTATTCAG-CGGGAG	
proBetl9	proBETL9 fwd	CGATGGTACTTACTCAT-GATGGTCATCTAGG	1911
	proBETL9 rev	CCATGGTATAACTTCAA-CTGTTGACGG	

Figure 7.14: The adventures of Marvin the Mad Maize.

References

Abrash, Emily B, & Bergmann, Dominique C. 2010. Regional specification of stomatal production by the putative ligand CHALLAH. *Development (Cambridge, England)*, **137**(3), 447–455.

Amien, Suseno, Kliwer, Irina, Márton, Mihaela L, Debener, Thomas, Geiger, Dietmar, Becker, Dirk, & Dresselhaus, Thomas. 2010. Defensin-like ZmES4 mediates pollen tube burst in maize via opening of the potassium channel KZM1. *PLoS biology*, **8**(6), e1000388.

Antoine, A F, Faure, J E, Cordeiro, S, Dumas, C, Rougier, M, & Feijo, J A. 2000. A calcium influx is triggered and propagates in the zygote as a wavefront during in vitro fertilization of flowering plants. *Proceedings of the National Academy of Sciences*, **97**(19), 10643–10648.

Antoine, A F, Faure, J E, Dumas, C, & Feijo, J A. 2001. Differential contribution of cytoplasmic Ca2+ and Ca2+ influx to gamete fusion and egg activation in maize. *Nature cell biology*, **3**(12), 1120–1123.

Balandín, Maite, Royo, Joaquín, Gómez, Elisa, Muñiz, Luis M, Molina, Antonio, & Hueros, Gregorio. 2005. A protective role for the embryo surrounding region of the maize endosperm, as evidenced by the characterisation of ZmESR-6, a defensin gene specifically expressed in this region. *Plant molecular biology*, **58**(2), 269–282.

Baneyx, François, & Mujacic, Mirna. 2004. Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotech*, **22**(11), 1399–1408.

Bate, Nicholas J, Niu, Xiping, Wang, Yuwen, Reimann, Kellie S, & Helentjaris, Timothy G. 2004. An invertase inhibitor from maize localizes to the embryo surrounding region during early kernel development. *Plant Physiology*, **134**(1), 246–254.

Bathgate, Ross A D, Zhang, Soude, Hughes, Richard A, Rosengren, K Johan, & Wade, John D. 2012. The structural determinants of insulin-like Peptide 3 activity. *Frontiers in Endocrinology*, **3**, 11.

Batoko, H, Zheng, H Q, Hawes, C, & Moore, I. 2000. A rab1 GTPase is required for transport between the endoplasmic reticulum and golgi apparatus and for normal golgi movement in plants. *Plant Cell*, **12**(11), 2201–2218.

Baud, Sébastien, Wuillème, Sylvie, Lemoine, Rémi, Kronenberger, Jocelyne, Caboche, Michel, Lepiniec, Loïc, & Rochat, Christine. 2005. The AtSUC5 sucrose transporter specifically expressed in the endosperm is involved in early seed development in Arabidopsis. *The Plant journal : for cell and molecular biology*, **43**(6), 824–836.

Baulcombe, D C. 1996. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant molecular biology*, **32**(1-2), 79–88.

Bayer, Martin, Nawy, Tal, Giglione, Carmela, Galli, Mary, Meinnel, Thierry, & Lukowitz, Wolfgang. 2009. Paternal control of embryonic patterning in Arabidopsis thaliana. *Science*, **323**(5920), 1485–1488.

Beddington, John. 2009. *Food, Energy, Water and the Climate: A Perfect Storm of Global Events?* Tech. rept.

Beets, Isabel, Janssen, Tom, Meelkop, Ellen, Temmerman, Liesbet, Suetens, Nick, Rademakers, Suzanne, Jansen, Gert, & Schoofs, Liliane. 2012. Vasopressin/oxytocin-related signaling regulates gustatory associative learning in C. elegans. *Science*, **338**(6106), 543–545.

Berger, Frédéric. 2008. Double-fertilization, from myths to reality. *Sexual Plant Reproduction*, **21**(1), 3–5.

Berger, Frédéric, Hamamura, Yuki, Ingouff, Mathieu, & Higashiyama, Tetsuya. 2008. Double fertilization - caught in the act. *Trends in plant science*, **13**(8), 437–443.

Bethke, PC, Lonsdale, JE, Fath, A, & Jones, RL. 1999. Hormonally regulated programmed cell death in barley aleurone cells. *THE PLANT CELL ONLINE*, **11**(6), 1033–1046.

Betsuyaku, Shigeyuki, Takahashi, Fuminori, Kinoshita, Atsuko, Miwa, Hiroki, Shinozaki, Kazuo, Fukuda, Hiroo, & Sawa, Shinichiro. 2011. Mitogen-activated protein kinase regulated by the CLAVATA receptors contributes to shoot apical meristem homeostasis. *Plant & cell physiology*, **52**(1), 14–29.

Blom, N, Gammeltoft, S, & Brunak, S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology*, **294**(11), 1351Ð1362.

Boch, Jens, Scholze, Heidi, Schornack, Sebastian, Landgraf, Angelika, Hahn, Simone, Kay, Sabine, Lahaye, Thomas, Nickstadt, Anja, & Bonas, Ulla. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**(5959), 1509–1512.

Boisson-Dernier, Aurélien, Roy, Sucharita, Kritsas, Konstantinos, Grobei, Monica A, Jaciubek, Miloslawa, Schroeder, Julian I, & Grossniklaus, Ueli. 2009. Disruption of the pollen-expressed FERONIA homologs ANXUR1 and ANXUR2 triggers pollen tube discharge. *Development (Cambridge, England)*, **136**(19), 3279–3288.

Bonello, J F, Opsahl-Ferstad, H G, Perez, P, Dumas, C, & Rogowsky, P M. 2000. Esr genes show different levels of expression in the same region of maize endosperm. *Gene*, **246**(1-2), 219–227.

Bonello, J F, Sevilla-Lecoq, S, Berne, A, Risueno, M C, Dumas, C, & Rogowsky, P M. 2002. Esr proteins are secreted by the cells of the embryo surrounding region. *J Exp Bot*, **53**(374), 1559–1568.

Brown, R C, Lemmon, B E, & Olsen, O A. 1994. Endosperm Development in Barley: Microtubule Involvement in the Morphogenetic Pathway. *THE PLANT CELL ONLINE*, **6**(9), 1241–1252.

Brown, R C, Lemmon, B E, & Olsen, O A. 1996a. Polarization predicts the pattern of cellularization in cereal endosperm. *Protoplasma*, **192**(3-4), 168–177.

Brown, R C, Lemmon, B E, & Nguyen, H. 2003. Events during the first four rounds of mitosis establish three developmental domains in the syncytial endosperm of Arabidopsis thaliana. *Protoplasma*, **222**(3-4), 167–174.

Brown, Roy C, Lemmon, Betty E, & Olsen, Odd-Arne. 1996b. Development of the endosperm in rice (Oryza sativa L.): Cellularization. *Journal of plant research*, **109**(3), 301–313.

Bruce, Toby J A. 2012. GM as a route for delivery of sustainable crop protection. *Journal of Experimental Botany*, **63**(2), 537–541.

Brummelkamp, Thijn R, Bernards, René, & Agami, Reuven. 2002. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer cell*, **2**(3), 243–247.

Bustin, Stephen A, Benes, Vladimir, Garson, Jeremy A, Hellemans, Jan, Huggett, Jim, Kubista, Mikael, Mueller, Reinhold, Nolan, Tania, Pfaffl, Michael W, Shipley, Gregory L, Vandesompele, Jo, & Wittwer, Carl T. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. **55**(4), 611Ð622.

Butenko, Melinka A, Vie, Ane Kjersti, Brembu, Tore, Aalen, Reidunn B, & Bones, Atle M. 2009. Plant peptides in signalling: looking for new partners. *Trends in plant science*, **14**(5), 255–263.

Carthew, R W. 2001. Gene silencing by double-stranded RNA. *Current opinion in cell biology*, **13**(2), 244–248.

Ceroni, Alessio, Passerini, Andrea, Vullo, Alessandro, & Frasconi, Paolo. 1. DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Research*, **34**(suppl 2), W177–W181.

Chae, Keun, Kieslich, Chris A, Morikis, Dimitrios, Kim, Seung-Chul, & Lord, Elizabeth M. 2009. A gain-of-function mutation of Arabidopsis lipid transfer protein 5 disturbs pollen tube tip growth and fertilization. *THE PLANT CELL ONLINE*, **21**(12), 3902–3914.

Chen, Junyi, Lausser, Andreas, & Dresselhaus, Thomas. 2014. Hormonal responses during early embryogenesis in maize. *Biochemical Society transactions*, **42**(2), 325–331.

Cheng, W H, Taliercio, E W, & Chourey, P S. 1996. The Miniature1 Seed Locus of Maize Encodes a Cell Wall Invertase Required for Normal Development of Endosperm and Maternal Cells in the Pedicel. *THE PLANT CELL ONLINE*, **8**(6), 971–983.

Chourey, P S, Jain, M, Li, Q B, & Carlson, S J. 2006. Genetic control of cell wall invertases in developing endosperm of maize. *Planta*, **223**, 159–167.

Clark, S E, Running, M P, & Meyerowitz, E M. 1995. CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. **121**, 2057–2067.

Clough, S J, & Bent, A F. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant journal : for cell and molecular biology*, **16**(6), 735–743.

Cobb, B G, & Hannah, L C. 1988. Shrunken-1 encoded sucrose synthase is not required for sucrose synthesis in the maize endosperm. *Plant Physiology*, **88**(4), 1219–1221.

Cole, Christian, Barber, Jonathan D, & Barton, Geoffrey J. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Research*, **36**(Web Server issue), W197–201.

Cook, D C, Fraser, R W, Paini, D R, Warden, A C, Lonsdale, W M, & De Barro, P J. 2011. Biosecurity and Yield Improvement Technologies Are Strategic Complements in the Fight against Food Insecurity. *PLoS One*, **6**(10), e26084.

Cordingley, M G, Callahan, P L, Sardana, V V, Garsky, V M, & Colonno, R J. 1990. Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J. Biol. Chem.*, **265**(16), 9062–9065.

Cossegal, M, Vernoud, V, Depege, N, & Rogowsky, P M. 2007. The Embryo Surrounding Region Endosperm. **7**, 57–71.

Costa, Liliana, Yuan, Jing, Rouster, Jacques, Paul, Wyatt, Dickinson, Hugh, & Gutierrez-Marcos, Jose. 2012. Maternal Control of Nutrient Allocation in Plant Seeds by Genomic Imprinting. *Current Biology*, **22**(2), 160–165.

Costa, Liliana M, Marshall, Eleanor, Tesfaye, Mesfin, Silverstein, Kevin A T, Mori, Masashi, Umetsu, Yoshitaka, Otterbach, Sophie L, Papareddy, Ranjith, Dickinson, Hugh G, Boutiller, Kim, VandenBosch, Kathryn A, Ohki, Shinya, & Gutierrez-Marcos, Jose F. 2014. Central cell-derived peptides regulate early embryo patterning in flowering plants. *Science*, **344**(6180), 168–172.

Curtis, Mark D, & Grossniklaus, Ueli. 2003. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiology*, **133**(2), 462–469.

Cutler, S R, Ehrhardt, D W, Griffitts, J S, & Somerville, C R. 2000. Random GFP::cDNA fusions enable visualization of subcellular structures in cells of Arabidopsis at a high frequency. *Proceedings of the National Academy of Sciences*, **97**(7), 3718–3723.

Davidson, R M, Hansey, C N, & Gowda, M. 2011. Utility of RNA sequencing for analysis of maize reproductive transcriptomes. *The Plant ...*, **4**(3), 191–203.

Davis, R W, Smith, J D, & Cobb, B G. 1990. A light and electron microscope investigation of the transfer cell region of maize caryopses. *Canadian Journal of Botany*, **68**(3), 471–479.

de Marco, Ario. 2009. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microbial Cell Factories*, **8**, 26.

De Smet, Ive, Voss, Ute, Jürgens, Gerd, & Beeckman, Tom. 2009. Receptor-like kinases shape the plant. *Nature cell biology*, **11**(10), 1166–1173.

Denoncin, Katleen, & Collet, Jean-Francois. 2013. Disulfide bond formation in the bacterial periplasm: major achievements and challenges ahead. *Antioxidants & redox signaling*, **19**(1), 63–71.

Diboll, A G. 1968. Fine structural development of the megagametophyte of Zea mays following fertilization. *American Journal of Botany*, **55**(7), 787–806.

Digonnet, C, Aldon, D, Leduc, N, Dumas, C, & Rougier, M. 1997. First evidence of a calcium transient in flowering plants at fertilization. *Development (Cambridge, England)*, **124**(15), 2867–2874.

Dohi, K, Nishikiori, M, Tamai, A, Ishikawa, M, Meshi, T, & Mori, M. 2006. Inducible virus-mediated expression of a foreign protein in suspension-cultured plant cells. *Archives of virology*, **151**(6), 1075–1084.

Dohi, Koji, & Mori, Masashi. 2007. Expression of active enzymes from an inducible tomato-mosaic-virus-based vector in cultured transgenic tobacco BY-2 cells. *Plant biotechnology*, **24**(4), 367–373.

Dohi, Koji, Tamai, Atsushi, & Mori, Masashi. 2008. Insertion in the coding region of the movement protein improves stability of the plasmid encoding a tomato mosaic virus-based expression vector. *Archives of virology*, **153**(9), 1667–1675.

Doran, Pauline M. 2006. Foreign protein degradation and instability in plants and plant tissue cultures. *Trends in Biotechnology*, **24**(9), 426–432.

Dresselhaus, Thomas, & Doughty, James. 2014. Regulation of fertilization and early seed development. *Biochemical Society transactions*, **42**(2), 309–312.

Elisabeth Gasteiger, Christine Hoogland Alexandre Gattiker S'everine Duvaud Marc Wilkins Ron Appel Amos Bairoch. 2005. *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press.

Ender, C, & Meister, G. 2010. Argonaute proteins at a glance. *Journal of cell science*, **123**, 1819–1823.

Engell, K. 1989. Embryology of barley: time course and analysis of controlled fertilization and early embryo formation based on serial sections. *Nordic journal of botany*, **9**(3), 265–280.

Escobar-Restrepo, Juan-Miguel, Huck, Norbert, Kessler, Sharon, Gagliardini, Valeria, Gheyselinck, Jacqueline, Yang, Wei-Cai, & Grossniklaus, Ueli. 2007. The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science*, **317**(5838), 656–660.

Fant, F, Vranken, W, Broekaert, W, & Borremans, F. 1998. Determination of the three-dimensional solution structure of Raphanus sativus antifungal protein 1 by 1H NMR. *Journal of molecular biology*, **279**(1), 257–270.

Feng, Wei, Pan, LiFeng, & Zhang, MingJie. 2011. Combination of NMR spectroscopy and X-ray crystallography offers unique advantages for elucidation of the structural basis of protein complex assembly. *Science China. Life sciences*, **54**(2), 101–111.

Finnegan, E J, Wang, M, & Waterhouse, P. 2001. Gene silencing: fleshing out the bones. *Curr Biol*, **11**(3), R99–R102.

Fischer, Markus. *Backtranslation Tool*.

Fischer, R, Vaquero-Martin, C, Sack, M, Drossard, J, Emans, N, & Commandeur, U. 1999. Towards molecular farming in the future: transient protein expression in plants. *Biotechnology and applied biochemistry*, **30 (Pt 2)**(Oct.), 113–116.

Flavell, R B. 1994. Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proceedings of the National Academy of Sciences of the United States of America*, **91**(9), 3490–3496.

Floyd, Sandra K, & Friedman, William E. 2000. Evolution of Endosperm Developmental Patterns among Basal Flowering Plants. *International Journal of Plant Sciences*, **161**(S6), S57–S81.

Forestan, Cristian, & Varotto, Serena. 2012. The role of PIN auxin efflux carriers in polar auxin transport and accumulation and their effect on shaping maize development. *Molecular plant*, **5**(4), 787–798.

Forestan, Cristian, Meda, Silvia, & Varotto, Serena. 2010. ZmPIN1-mediated auxin transport is related to cellular differentiation during maize embryogenesis and endosperm development. *Plant Physiology*, **152**(3), 1373–1390.

Gao, R, Dong, S, Fan, J, & Hu, C. 1998. Relationship between development of endosperm transfer cells and grain mass in maize. *Biologia plantarum*, **123**, 1819–1823.

Giuliani, Concetta, Consonni, Gabriella, Gavazzi, Giuseppe, Colombo, Monica, & Dolfini, Silvana. 2002. Programmed cell death during embryogenesis in maize. *Annals of botany*, **90**(2), 287–292.

Godfray, H C, Beddington, J R, Crute, I R, Haddad, L, Lawrence, D, Muir, J F, Pretty, J, Robinson, S, Thomas, S M, & Toulmin, C. 2010. Food security: the challenge of feeding 9 billion people. *Science*, **327**(5967), 812–818.

Gómez, Elisa, Royo, Joaquín, Guo, Yan, Thompson, Richard, & Hueros, Gregorio. 2002. Establishment of cereal endosperm expression domains: identification and properties of a maize transfer cell-specific transcription factor, ZmMRP-1. *Plant Cell*, **14**(3), 599–610.

Gómez, Elisa, Royo, Joaquín, Muñiz, Luis M, Sellam, Olivier, Paul, Wyatt, Gerentes, Denise, Barrero, Cristina, López, Maribel, Perez, Pascual, & Hueros, Gregorio. 2009. The maize transcription factor myb-related protein-1 is a key regulator of the differentiation of transfer cells. *Plant Cell*, **21**(7), 2022–2035.

Graham, Michelle A, Silverstein, Kevin A T, Cannon, Steven B, & VandenBosch, Kathryn A. 2004. Computational identification and characterization of novel genes from legumes. *Plant Physiology*, **135**(3), 1179–1197.

Greenbaum, Dov, Jansen, Ronald, & Gerstein, Mark. 2002. Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics*, **18**(4), 585–596.

Grignon, C, & Sentenac, H. 1991. pH and ionic conditions in the apoplast. *Annual review of plant biology*, **42**, 103–128.

Gruis, Darren Fred, Guo, Hena, Selinger, David, Tian, Qing, & Olsen, Odd-Arne. 2006. Surface position, not signaling from surrounding maternal tissues, specifies aleurone epidermal cell fate in maize. *Plant Physiology*, **141**(3), 898–909.

Gry, Marcus, Rimini, Rebecca, Strömberg, Sara, Asplund, Anna, Pontén, Fredrik, Uhlén, Mathias, & Nilsson, Peter. 2009. Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC genomics*, **10**, 365.

Guignard, M L. 1899. Sur les antherozoides et la double copulation sexuelle chez les vegetaux angiosperms. *Rev Gén Bot*, **11**, 129–135.

Gutierrez-Marcos, J F, Costa, L M, Biderre-Petit, C, Khbaya, B, O'Sullivan, D M, Wormald, M, Perez, P, & Dickinson, H G. 2004. maternally expressed gene1 Is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. *Plant Cell*, **16**(5), 1288–1301.

Gygi, S P, Rochon, Y, Franza, B R, & Aebersold, R. 1999. Correlation between protein and mRNA abundance in yeast. *Molecular and cellular biology*, **19**(3), 1720–1730.

Hajduch, M, Khaled, A, Depege, N, & Rogowsky, M P. 2005. Maize embryogenesis. *Maydica*, **50**, 469–483.

Hamamura, Yuki, Nishimaki, Moe, Takeuchi, Hidenori, Geitmann, Anja, Kurihara, Daisuke, & Higashiyama, Tetsuya. 2014. Live imaging of calcium spikes during double fertilization in Arabidopsis. *Nature Communications*, **5**, 4722.

Hammami, Riadh, Ben Hamida, Jeannette, Vergoten, Gérard, & Fliss, Ismail. 2009. PhytAMP: a database dedicated to antimicrobial plant peptides. *Nucleic Acids Research*, **37**(Database issue), D963–968.

Hara, Kenta, Kajita, Ryoko, Torii, Keiko U, Bergmann, Dominique C, & Kakimoto, Tatsuo. 2007. The secretory peptide gene EPF1 enforces the stomatal one-cell-spacing rule. *Genes Dev*, **21**(14), 1720–1725.

Hara, Kenta, Yokoo, Toshiya, Kajita, Ryoko, Onishi, Takaaki, Yahata, Saiko, Peterson, Kylee M, Torii, Keiko U, & Kakimoto, Tatsuo. 2009. Epidermal cell density is autoregulated via a secretory peptide, EPIDERMAL PATTERNING FACTOR 2 in Arabidopsis leaves. *Plant & cell physiology*, **50**(6), 1019–1031.

Haruta, Miyoshi, Sabat, Grzegorz, Stecker, Kelly, Minkoff, Benjamin B, & Sussman, Michael R. 2014. A peptide hormone and its receptor protein kinase regulate plant cell expansion. *Science*, **343**(6169), 408–411.

Hatahet, Feras, Nguyen, Van Dat, Salo, Kirsi, & Ruddock, Lloyd. 2010. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microbial Cell Factories*, **9**(1), 67–67.

Hellwig, Stephan, Drossard, Jurgen, Twyman, Richard M, & Fischer, Rainer. 2004. Plant cell cultures for the production of recombinant proteins. *Nat Biotech*, **22**(11), 1415–1422.

Higashiyama, T, Yabe, S, Sasaki, N, Nishimura, Y, S, Miyagishima, Kuroiwa, H, & Kuroiwa, T. 2001. Pollen tube attraction by the synergid cell. *Science*, **293**(5534), 1480–1483.

Huang, X, & Miller, W. 1991. A time-efficient, linear-space local similarity algorithm. *Advances in Applied Mathematics*, **12**(3), 337–357.

Hueros, G, Varotto, S, Salamini, F, & Thompson, R D. 1995. Molecular characterization of BET1, a gene expressed in the endosperm transfer cells of maize. *Plant Cell*, **7**(6), 747–757.

Hueros, G, Gomez, E, Cheikh, N, Edwards, J, Weldon, M, Salamini, F, & Thompson, RD. 1999a. Identification of a promoter sequence from the BETL1 gene cluster able to confer transfer-cell-specific expression in transgenic maize. *Plant Physiology*, **121**(4), 1143–1152.

Hueros, Gregorio, Royo, Joaquín, Maitz, Monika, Salamini, Francesco, & Thompson, Richard D. 1999b. Evidence for factors regulating transfer cell-specific expression in maize endosperm. *Plant molecular biology*, **41**(3), 403–414.

Hunt, Lee, & Gray, Julie E. 2009. The signaling peptide EPF2 controls asymmetric cell divisions during stomatal development. *Curr Biol*, **19**(10), 864–869.

Ingram, G C, Magnard, J L, Vergne, P, Dumas, C, & Rogowsky, P M. 1999. ZmOCL1, an HDGL2 family homeobox gene, is expressed in the outer cell layer throughout maize development. *Plant molecular biology*, **40**(2), 343–354.

Ingram, G C, Boisnard-Lorig, C, Dumas, C, & Rogowsky, P M. 2000. Expression patterns of genes encoding HD-ZipIV homeo domain proteins define specific domains in maize embryos and meristems. *The Plant journal : for cell and molecular biology*, **22**(5), 401–414.

Ishida, Y, Saito, H, Ohta, S, Hiei, Y, Komari, T, & Kumashiro, T. 1996. High efficiency transformation of maize (Zea mays L.) mediated by Agrobacterium tumefaciens. *Nat Biotech*, **14**(6), 745–750.

Ishimaru, Ken, Hirotsu, Naoki, Madoka, Yuka, Murakami, Naomi, Hara, Nao, Onodera, Haruko, Kashiwagi, Takayuki, Ujiie, Kazuhiro, Shimizu, Bun-Ichi, Onishi, Atsuko, Miyagawa, Hisashi, & Katoh, Etsuko. 2013. Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nature genetics*, **45**(6), 707–711.

Jacobsen, J V, Knox, R B, & Pyliotis, N A. 1971. The structure and composition of aleurone grains in the barley aleurone layer. *Planta*, **101**(3), 189–209.

Jefferson, R A, Burgess, S M, & Hirsh, D. 1986. beta-Glucuronidase from Escherichia coli as a gene-fusion marker. *Proceedings of the National Academy of Sciences of the United States of America*, **83**(22), 8447–8451.

Jefferson, R A, Kavanagh, T A, & Bevan, M W. 1987a. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO journal*, **6**(13), 3901–3907.

Jefferson, R A, Bevan, M, & Kavanagh, T. 1987b. The use of the Escherichia coli beta-glucuronidase as a gene fusion marker for studies of gene expression in higher plants. *Biochemical Society transactions*, **15**(1), 17–18.

Jewaria, Pawan Kumar, Hara, Toshiaki, Tanaka, Hirokazu, Kondo, Tatsuhiko, Betsuyaku, Shigeyuki, Sawa, Shinichiro, Sakagami, Youji, Aimoto, Saburo, & Kakimoto, Tatsuo. 2013. Differential effects of the peptides Stomagen, EPF1 and EPF2 on activation of MAP kinase MPK6 and the SPCH protein level. *Plant & cell physiology*, **54**(8), 1253–1262.

Jones, J D. 2011. Why genetically modified crops? *Philos Transact A Math Phys Eng Sci*, **369**(1942), 1807–1816.

Jones, R L. 1969. The fine structure of barley aleurone cells. *Planta*, **85**(4), 359–375.

Kawashima, Tomokazu, & Goldberg, Robert B. 2010. The suspensor: not just suspending the embryo. *Trends in plant science*, **15**(1), 23–30.

Kelly, Sharon M, Jess, Thomas J, & Price, Nicholas C. 2005. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1751**(2), 119–139.

Kiesselbach, T A, & Walker, E R. 1952. Structure of certain specialized tissue in the kernel of corn. *American Journal of Botany*, **39**(8), 561–570.

Kim, J Y, Mahé, A, Guy, S, Brangeon, J, Roche, O, Chourey, P S, & Prioul, J L. 2000. Characterization of two members of the maize gene family, Incw3 and Incw4, encoding cell-wall invertases. *Gene*, **245**(1), 89–102.

Kiyosue, T, Ohad, N, Yadegari, R, Hannon, M, Dinneny, J, Wells, D, Katz, A, Margossian, L, Harada, J J, Goldberg, R B, & Fischer, R L. 1999. Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **96**(7), 4186–4191.

Kladnik, Ales, Chamusco, Karen, Dermastia, Marina, & Chourey, Prem. 2004. Evidence of programmed cell death in post-phloem transport cells of the maternal pedicel tissue in developing caryopsis of maize. *Plant Physiology*, **136**(3), 3572–3581.

Komari, T, Hiei, Y, Saito, Y, Murai, N, & Kumashiro, T. 1996. Vectors carrying two separate T-DNAs for co-transformation of higher plants mediated by Agrobacterium tumefaciens and segregation of transformants free from selection markers. *The Plant journal : for cell and molecular biology*, **10**(1), 165–174.

Koncz, Csaba, & Schell, Jeff. 1986. The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Molecular and General Genetics MGG*, **204**(3), 383–396.

Kondo, Tatsuhiko, Kajita, Ryoko, Miyazaki, Aya, Hokoyama, Mayumi, Nakamura-Miura, Touko, Mizuno, Satoko, Masuda, Yuichi, Irie, Kazuhiro, Tanaka, Yuki, Takada, Shinobu, Kakimoto, Tatsuo, & Sakagami, Youji. 2010. Stomatal density is controlled by a mesophyll-derived signaling molecule. *Plant & cell physiology*, **51**(1), 1–8.

Kordyum, E L. 2008. Double fertilization in flowering plants: 1898–2008. *Cytology and Genetics*, **42**(3), 147–158.

Kozieradzka-Kiszkurno, Małgorzata, & Płachno, Bartosz Jan. 2011. Are there symplastic connections between the endosperm and embryo in some angiosperms?—a lesson from the Crassulaceae family. *Protoplasma*, **249**(4), 1081–1089.

Kranz, E, von Wiegen P, Quader, H, & Lorz, H. 1998. Endosperm development after fusion of isolated, single maize sperm and central cells in vitro. *THE PLANT CELL ONLINE*, **10**(4), 511–524.

Kurup, Smita, Runions, John, Köhler, Uwe, Laplaze, Laurent, Hodge, Sarah, & Haseloff, Jim. 2005. Marking cell lineages in living tissues. *The Plant journal : for cell and molecular biology*, **42**(3), 444–453.

Kyle, D J, & Styles, E D. 1977. Development of aleurone and sub-aleurone layers in maize. *Planta*, **137**(3), 185–193.

Laemmli. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**(5259), 680–685.

Lakatos, Lóránt, Szittya, György, Silhavy, Dániel, & Burgyán, József. 2004. Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. *The EMBO journal*, **23**(4), 876–884.

Lammeren, A A. 1987. Embryogenesis in Zea mays L.: a structural approach to maize caryopsis development in vivo and in vitro.

Le, Quyen, Gutierrez-Marcos, Jose F, Costa, Liliana M, Meyer, Stephanie, Dickinson, Hugh G, Lorz, Horst, Kranz, Erhard, & Scholten, Stefan. 2005. Construction and screening of subtracted cDNA libraries from limited populations of plant cells: a comparative analysis of gene expression between maize egg cells and central cells. *The Plant Journal*, **44**(1), 167–178.

LeCLere, Sherry, Schmelz, Eric A, & Chourey, Prem S. 2010. Sugar levels regulate tryptophan-dependent auxin biosynthesis in developing maize kernels. *Plant Physiology*, **153**(1), 306–318.

Leduc, N, Matthys-Rochon, E, Rougier, M, Mogensen, L, Holm, P, Magnard, J L, & Dumas, C. 1996. Isolated maize zygotes mimic in vivo embryonic development and

express microinjected genes when cultured in vitro. *Developmental biology*, **177**(1), 190–203.

Lees, Jonathan G, Miles, Andrew J, Wien, Frank, & Wallace, B A. 2006. A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics*, **22**(16), 1955–1962.

Liang, Zhen, Zhang, Kang, Chen, Kunling, & Gao, Caixia. 2014. Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. *Journal of genetics and genomics = Yi chuan xue bao*, **41**(2), 63–68.

Liu, Jian-Xiang, Srivastava, Renu, & Howell, Stephen. 2009. Overexpression of an Arabidopsis gene encoding a subtilase (AtSBT5.4) produces a clavata-like phenotype. *Planta*, **230**(4), 687–697.

Lobley, A, Whitmore, L, & Wallace, B A. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*, **18**(1), 211–212.

Lobstein, J, Emrich, C A, Jeans, C, & Faulkner, M. 2012. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell ...*, **11**(56), 1–16.

Lohaus, Gertrud, Pennewiss, Kerstin, Sattelmacher, Burkhard, Hussmann, Melanie, & Hermann Muehling, Karl. 2001. Is the infiltration-centrifugation technique appropriate for the isolation of apoplastic fluid? A critical evaluation with different plant species. *Physiologia Plantarum*, **111**(4), 457–465.

Lopes, M A, & Larkins, B A. 1993. Endosperm origin, development, and function. *Plant Cell*, **5**(10), 1383–1399.

Lukowitz, Wolfgang, Roeder, Adrienne, Parmenter, Dana, & Somerville, Chris. 2004. A MAPKK kinase gene regulates extra-embryonic cell fate in Arabidopsis. *Cell*, **116**(1), 109–119.

Luo, Ming, Dennis, Elizabeth S, Berger, Frédéric, Peacock, William James, & Chaudhury, Abed. 2005. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(48), 17531–17536.

Magnard, J L, Le Deunff, E, Domenech, J, Rogowsky, P M, Testillano, P S, Rougier, M, Risueno, M C, Vergne, P, & Dumas, C. 2000. Genes normally expressed in the endosperm are expressed at early stages of microspore embryogenesis in maize. *Plant molecular biology*, **44**(4), 559–574.

Magnard, Jean-Louis, Lehouque, Gaëlle, Massonneau, Agnès, Frangne, Nathalie, Heckel, Thierry, Gutierrez-Marcos, Jose F, Perez, Pascual, Dumas, Christian, & Rogowsky, Peter M. 2003. ZmEBE genes show a novel, continuous expression pattern in the central cell before fertilization and in specific domains of the resulting endosperm after fertilization. *Plant molecular biology*, **53**(6), 821–836.

Mansfield, S G, & Briarty, L G. 1990. Development of the free-nuclear endosperm in Arabidopsis thaliana (L.). *Arabidopsis Inf Serv*, **27**, 1–8.

Marshall, E, Costa, L M, & Gutierrez-Marcos, J. 2011. Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J Exp Bot*, **62**(5), 1677–1686.

Márton, Mihaela L, Cordts, Simone, Broadhvest, Jean, & Dresselhaus, Thomas. 2005. Micropylar pollen tube guidance by egg apparatus 1 of maize. *Science*, **307**(5709), 573–576.

Matzke, Marjori A, Aufsatz, Werner, Kanno, Tatsuo, Mette, M Florian, & Matzke, Antonius J M. 2002. Homology-dependent gene silencing and host defense in plants. *Advances in genetics*, **46**, 235–275.

McCarty, Donald R, Settles, Andrew Mark, Suzuki, Masaharu, Tan, Bao Cai, Latshaw, Susan, Porch, Tim, Robin, Kevin, Baier, John, Avigne, Wayne, Lai, Jinsheng, Messing, Joachim, Koch, Karen E, & Hannah, L Curtis. 2005. Steady-state transposon mutagenesis in inbred maize. *The Plant journal : for cell and molecular biology*, **44**(1), 52–61.

Mergaert, Peter, Nikovics, Krisztina, Kelemen, Zsolt, Maunoury, Nicolas, Vaubert, Danièle, Kondorosi, Adam, & Kondorosi, Eva. 2003. A novel family in Medicago truncatula consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiology*, **132**(1), 161–173.

Miller, M E, & Chourey, P S. 1992. The Maize Invertase-Deficient miniature-1 Seed Mutation Is Associated with Aberrant Pedicel and Endosperm Development. *THE PLANT CELL ONLINE*, **4**(3), 297–305.

Miller, Victor M, Xia, Haibin, Marrs, Ginger L, Gouvion, Cynthia M, Lee, Gloria, Davidson, Beverly L, & Paulson, Henry L. 2003. Allele-specific silencing of dominant disease genes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(12), 7195–7200.

Mingossi, Fabiana B, Matos, Juliana L, Rizzato, Ana Paula, Medeiros, Ane H, Falco, Maria C, Silva-Filho, Marcio C, & Moura, Daniel S. 2010. SacRALF1, a peptide signal from the grass sugarcane (Saccharum spp.), is potentially involved in the regulation of tissue expansion. *Plant molecular biology*, **73**(3), 271–281.

Mishima, Masaki, Takayama, Seiji, Sasaki, Kei-Ichi, Jee, Jun-Goo, Kojima, Chojiro, Isogai, Akira, & Shirakawa, Masahiro. 2003. Structure of the male determinant factor for Brassica self-incompatibility. *J. Biol. Chem.*, **278**(38), 36389–36395.

Miyazaki, Saori, Murata, Takashi, Sakurai-Ozato, Nami, Kubo, Minoru, Demura, Taku, Fukuda, Hiroo, & Hasebe, Mitsuyasu. 2009. ANXUR1 and 2, sister genes to FERONIA/SIRENE, are male factors for coordinated fertilization. *Curr Biol*, **19**(15), 1327–1331.

Mòl, R, Matthys-Rochon, E, & Dumas, C. 1993. In-vitro culture of fertilized embryo sacs of maize: zygotes and two-celled proembryos can develop into plants. *Planta*, **189**(2), 213–217.

Mòl, R, Rochon, E Matthys, & Dumas, C. 1994. The kinetics of cytological events during double fertilization in Zea mays L. *The Plant Journal*, **5**(2), 197–206.

Mollet, J C, Park, S Y, Nothnagel, E A, & Lord, E M. 2000. A lily stylar pectin is necessary for pollen tube adhesion to an in vitro stylar matrix. *Plant Cell*, **12**(9), 1737–1750.

Moore, Ian, Samalova, Marketa, & Kurup, Smita. 2006. Transactivated and chemically inducible gene expression in plants. *The Plant journal : for cell and molecular biology*, **45**(4), 651–683.

Morgenstern, Burkhard, Prohaska, Sonja J, Pöhler, Dirk, & Stadler, Peter F. 2006. Multiple sequence alignment with user-defined anchor points. *Algorithms for molecular biology : AMB*, **1**(1), 6.

Morrison, I N, Kuo, J, & O'brien, T P. 1975. Histochemistry and fine structure of developing wheat aleurone cells. *Planta*, **123**(2), 105–116.

Mouillon, Jean-Marie, Gustafsson, Petter, & Harryson, Pia. 2006. Structural Investigation of Disordered Stress Proteins. Comparison of Full-Length Dehydrins with Isolated Peptides of Their Conserved Segments. *Plant Physiology*, **141**(2), 638–650.

Muñiz, Luis M, Royo, Joaquín, Gómez, Elisa, Barrero, Cristina, Bergareche, Diego, & Hueros, Gregorio. 2006. The maize transfer cell-specific type-A response regulator ZmTCRR-1 appears to be involved in intercellular signalling. *The Plant journal : for cell and molecular biology*, **48**(1), 17–27.

Müntz, K. 1998. Deposition of storage proteins. *Protein Trafficking in Plant Cells*, **1**, 77–99.

Murase, Kohji, Shiba, Hiroshi, Iwano, Megumi, Che, Fang-Sik, Watanabe, Masao, Isogai, Akira, & Takayama, Seiji. 2004. A membrane-anchored protein kinase involved in Brassica self-incompatibility signaling. *Science*, **303**(5663), 1516–1519.

Muschietti, J, Dircks, L, Vancanneyt, G, & McCormick, S. 1994. LAT52 protein is essential for tomato pollen development: pollen expressing antisense LAT52 RNA hydrates and germinates abnormally and cannot achieve fertilization. *The Plant journal : for cell and molecular biology*, **6**(3), 321–338.

Mussolino, Claudio, & Cathomen, Toni. 2013. RNA guides genome engineering. *Nature biotechnology*, **31**(3), 208–209.

Nadeau, Jeanette A, & Sack, Fred D. 2002. Control of stomatal distribution on the Arabidopsis leaf surface. *Science*, **296**(5573), 1697–1700.

Nagalakshmi, Ugrappa, Wang, Zhong, Waern, Karl, Shou, Chong, Raha, Debasish, Gerstein, Mark, & Snyder, Michael. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, **320**(5881), 1344–1349.

Nagasawa, Nobuhiro, Hibara, Ken-ichiro, Heppard, Elmer P, Vander Velden, Kent A, Luck, Stanley, Beatty, Mary, Nagato, Yasuo, & Sakai, Hajime. 2013. GIANT EMBRYO encodes CYP78A13, required for proper size balance between embryo and endosperm in rice. *The Plant journal : for cell and molecular biology*, **75**(4), 592–605.

Nagl, Walter. 1990. Translocation of Putrescine in the Ovule, Suspensor and Embryo of Phaseolus coccineus. *Journal of Plant Physiology*, **136**(5), 587–591.

Nguyen, H, Brown, R C, & Lemmon, B E. 2001. Patterns of Cytoskeletal Organization Reflect Distinct Developmental Domains in Endosperm of Coronopus didymus(Brassicaceae). *International Journal of Plant Sciences*, **162**(1), 1–14.

Nguyen, Van Dat, Hatahet, Feras, Salo, Kirsi E H, Enlund, Eveliina, Zhang, Chi, & Ruddock, Lloyd W. 2011. Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. *Microbial Cell Factories*, **10**, 1.

Norton, R S, & Pallaghy, P K. 1998. The cystine knot structure of ion channel toxins and related polypeptides. *Toxicon : official journal of the International Society on Toxinology*, **36**(11), 1573–1583.

Nowack, Moritz K, Shirzadi, Reza, Dissmeyer, Nico, Dolf, Andreas, Endl, Elmar, Grini, Paul E, & Schnittger, Arp. 2007. Bypassing genomic imprinting allows seed development. *Nature*, **447**(7142), 312–315.

Offler, Christina E, McCurdy, David W, Patrick, John W, & Talbot, Mark J. 2003. Transfer cells: cells specialized for a special purpose. *Annual review of plant biology*, **54**, 431–454.

Ofran, Yanay, & Rost, Burkhard. 2007. ISIS: interaction sites identified from sequence. *Bioinformatics*, **23**(2), e13–e16.

Ohad, N, Yadegari, R, Margossian, L, Hannon, M, Michaeli, D, Harada, J J, Goldberg, R B, & Fischer, R L. 1999. Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. *Plant Cell*, **11**(3), 407–416.

Ohki, Shinya, Dohi, Koji, Tamai, Atsushi, Takeuchi, Makoto, & Mori, Masashi. 2008. Stable-isotope labeling using an inducible viral infection system in suspension-cultured plant cells. *Journal of biomolecular NMR*, **42**(4), 271–277.

Ohki, Shinya, Takeuchi, Makoto, & Mori, Masashi. 2011. The NMR structure of stomagen reveals the basis of stomatal density regulation by plant peptide hormones. *Nat Commun*, **2**, 512–512.

Okuda, Satohiro, Tsutsui, Hiroki, Shiina, Keiko, Sprunck, Stefanie, Takeuchi, Hidenori, Yui, Ryoko, Kasahara, Ryushiro D, Hamamura, Yuki, Mizukami, Akane, Susaki, Daichi, Kawano, Nao, Sakakibara, Takashi, Namiki, Shoko, Itoh, Kie, Otsuka, Kurataka, Matsuzaki, Motomichi, Nozaki, Hisayoshi, Kuroiwa, Tsuneyoshi, Nakano, Akihiko, Kanaoka, Masahiro M, Dresselhaus, Thomas, Sasaki, Narie, & Higashiyama, Tetsuya. 2009. Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature*, **458**(7236), 357–361.

Olsen, Addie Nina, Mundy, John, & Skriver, Karen. 2002. Peptomics, identification of novel cationic Arabidopsis peptides with conserved sequence motifs. *In silico biology*, **2**(4), 441–451.

Olsen, O A. 2004. Dynamics of maize aleurone cell formation: the" surface" rule. *Maydica*, **49**, 37–40.

Olsen, O A, Brown, R C, & Lemmon, B E. 1995. Pattern and process of wall formation in developing endosperm. *BioEssays*, **17**(9), 803–812.

Olsen, Odd-Arne. 2007. *Endosperm*. Developmental and Molecular Biology. Springer Science & Business Media.

Opsahl-Ferstad, H G, Le Deunff, E, Dumas, C, & Rogowsky, P M. 1997. ZmEsr, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. *The Plant journal : for cell and molecular biology*, **12**(1), 235–246.

Pallaghy, P K, Nielsen, K J, Craik, D J, & Norton, R S. 1994. A common structural motif incorporating a cystine knot and a triple-stranded beta-sheet in toxic and inhibitory polypeptides. *Protein science : a publication of the Protein Society*, **3**(10), 1833–1839.

Pare, J M, & Hobman, T C. 2007. Dicer: Structure, Function And Role In RNA-Dependent Gene-Silencing Pathways. *Industrial Enzymes*, **1**, 421–438.

Pate, J S, & Gunning, B E S. 1972. Transfer Cells. *Annual Review of Plant Physiology*, **23**(1), 173–196.

Pearce, G, Moura, D S, Stratmann, J, & Ryan, C A. 2001. RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(22), 12843–12847.

Petersen, Thomas Nordahl, Brunak, Soren, von Heijne, Gunnar, & Nielsen, Henrik. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth*, **8**(10), 785–786.

Pfaffl, M W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, **29**(9), e45.

Puchta, Holger, & Fauser, Friedrich. 2013. Gene targeting in plants: 25 years later. *The International journal of developmental biology*, **57**(6-8), 629–637.

Raghavan, V. 2006. *Double fertilization: embryo and endosperm development in flowering plants*. Springer Berlin / Heidelberg.

Redfern, A D, Colley, S M, & Beveridge, D J. 2013. RNA-induced silencing complex (RISC) Proteins PACT, TRBP, and Dicer are SRA binding nuclear receptor coregulators. *Proceedings of the National Academy of Sciences*, **110**(16), 6536Ð–6541.

Richardson, Lynn G L, & Torii, Keiko U. 2013. Take a deep breath: peptide signalling in stomatal patterning and differentiation. *J Exp Bot*, Aug.

Riley, Kasandra J, Yario, Therese A, & Steitz, Joan A. 2012. Association of Argonaute proteins and microRNAs can occur after cell lysis. *RNA (New York, N.Y.)*, **18**(9), 1581–1585.

Rost, Burkhard, Yachdav, Guy, & Liu, Jinfeng. 2004. The PredictProtein server. *Nucleic Acids Research*, **32**(suppl 2), W321–W326.

Saaranen, Mirva J, & Ruddock, Lloyd W. 2012. Disulfide Bond Formation in the Cytoplasm. *Antioxidants & redox signaling*, **00**(Oct.), 1–8.

Samalova, Marketa, Brzobohaty, Bretislav, & Moore, Ian. 2005. pOp6/LhGR: a stringently regulated and highly responsive dexamethasone-inducible gene expression system for tobacco. *The Plant journal : for cell and molecular biology*, **41**(6), 919–935.

Sander, A, Krausgrill, S, Greiner, S, Weil, M, & Rausch, T. 1996. Sucrose protects cell wall invertase but not vacuolar invertase against proteinaceous inhibitors. *FEBS letters*, **385**(3), 171–175.

Santella, L, Lim, D, & Moccia, F. 2004. Calcium and fertilization: the beginning of life. *Trends in biochemical sciences*, **29**(8), 401–408.

Saurabh, Satyajit, Vidyarthi, Ambarish S, & Prasad, Dinesh. 2014. RNA interference: concept to reality in crop improvement. *Planta*, **239**(3), 543–564.

Schägger, Hermann. 2006. Tricine-SDS-PAGE. *Nature protocols*, **1**(1), 16–22.

Schel, J H N, Kieft, H, & Lammeren, A A M Van. 1984. Interactions between embryo and endosperm during early developmental stages of maize caryopses (Zea mays). *Can. J. Bot.*, **62**(12), 2842–2853.

Scheres, B, van Engelen, F, van der Knaap, E, Van De Wiel, C, Van Kammen, A, & Bisseling, T. 1990a. Sequential induction of nodulin gene expression in the developing pea nodule. *Plant Cell*, **2**(8), 687–700.

Scheres, B, Van De Wiel, C, Zalensky, A, Horvath, B, Spaink, H, Van Eck, H, Zwartkruis, F, Wolters, A M, Gloudemans, T, & Van Kammen, A. 1990b. The ENOD12 gene product is involved in the infection process during the pea-Rhizobium interaction. *Cell*, **60**(2), 281–294.

Schlessinger, Avner, & Rost, Burkhard. 2005. Protein flexibility and rigidity predicted from sequence. *Proteins*, **61**(1), 115–126.

Schlessinger, Avner, Yachdav, Guy, & Rost, Burkhard. 2006. PROFbval: predict flexible and rigid residues in proteins. *Bioinformatics*, **22**(7), 891–893.

Schlessinger, Avner, Punta, Marco, Yachdav, Guy, Kajan, Laszlo, & Rost, Burkhard. 2009. Improved Disorder Prediction by Combination of Orthogonal Approaches. *PLoS One*, **4**(2), e4433–.

Schopfer, C R, Nasrallah, M E, & Nasrallah, J B. 1999. The male determinant of self-incompatibility in Brassica. *Science*, **286**(5445), 1697–1700.

Sekhon, Rajandeep S, Lin, Haining, Childs, Kevin L, Hansey, Candice N, Buell, C Robin, de Leon, Natalia, & Kaeppler, Shawn M. 2011. Genome-wide atlas of transcription during maize development. *The Plant journal : for cell and molecular biology*, **66**(4), 553–563.

Serna, A, Maitz, M, O'Connell, T, Santandrea, G, Thevissen, K, Tienens, K, Hueros, G, Faleri, C, Cai, G, Lottspeich, F, & Thompson, R D. 2001. Maize endosperm secretes a novel antifungal protein into adjacent maternal tissue. *The Plant journal : for cell and molecular biology*, **25**(6), 687–698.

Sevilla-Lecoq, Sandrine, Deguerry, Fabienne, Matthys-Rochon, Elisabeth, Perez, Pascual, Dumas, Christian, & Rogowsky, Peter M. 2003. Analysis of ZmAE3 upstream sequences in maize endosperm and androgenic embryos. *Sexual Plant Reproduction*, **16**(1), 1–8.

Shih, Sharon M H, & Doran, Pauline M. 2009. Foreign protein production using plant cell and organ cultures: Advantages and limitations. *Biotechnology Advances*, **27**(6), 1036–1042.

Shpak, Elena D, McAbee, Jessica Messmer, Pillitteri, Lynn Jo, & Torii, Keiko U. 2005. Stomatal patterning and differentiation by synergistic interactions of receptor kinases. *Science*, **309**(5732), 290–293.

Silverstein, Kevin A T, Moskal, William A, Wu, Hank C, Underwood, Beverly A, Graham, Michelle A, Town, Christopher D, & VandenBosch, Kathryn A. 2007. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *The Plant journal : for cell and molecular biology*, **51**(2), 262–280.

Smart, M G, & O'brien, T P. 1983. The development of the wheat embryo in relation to the neighbouring tissues. *Protoplasma*, **114**(1-2), 1–13.

Sørensen, M B, Chaudhury, A M, Robert, H, Bancharel, E, & Berger, F. 2001. Polycomb group genes control pattern formation in plant seed. *Curr Biol*, **11**(4), 277–281.

Sossountzov, L, Ruiz-Avila, L, Vignols, F, Jolliot, A, Arondel, V, Tchang, F, Grosbois, M, Guerbette, F, Miginiac, E, & Delseny, M. 1991. Spatial and temporal expression of a maize lipid transfer protein gene. *Plant Cell*, **3**(9), 923–933.

Sprunck, Stefanie, Rademacher, Svenja, Vogler, Frank, Gheyselinck, Jacqueline, Grossniklaus, Ueli, & Dresselhaus, Thomas. 2012. Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science*, **338**(6110), 1093–1097.

Sreerama, Narasimha, & Woody, Robert W. 2000. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and

CDSSTR Methods with an Expanded Reference Set. *Analytical Biochemistry*, **287**(2), 252–260.

Stadler, Ruth, Lauterbach, Christian, & Sauer, Norbert. 2005. Cell-to-cell movement of green fluorescent protein reveals post-phloem transport in the outer integument and identifies symplastic domains in Arabidopsis seeds and embryos. *Plant Physiology*, **139**(2), 701–712.

Stein, J C, Howlett, B, Boyes, D C, Nasrallah, M E, & Nasrallah, J B. 1991. Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of Brassica oleracea. *Proceedings of the National Academy of Sciences of the United States of America*, **88**(19), 8816–8820.

Stewart, A, Nield, H, & Lott, J N. 1988. An investigation of the mineral content of barley grains and seedlings. *Plant Physiology*, **86**(1), 93–97.

Sugano, Shigeo S, Shimada, Tomoo, Imai, Yu, Okawa, Katsuya, Tamai, Atsushi, Mori, Masashi, & Hara-Nishimura, Ikuko. 2010. Stomagen positively regulates stomatal density in Arabidopsis. *Nature*, **463**(7278), 241–244.

Suzuki, G, Kai, N, Hirose, T, Fukui, K, Nishio, T, Takayama, S, Isogai, A, Watanabe, M, & Hinata, K. 1999. Genomic organization of the S locus: Identification and characterization of genes in SLG/SRK region of S(9) haplotype of Brassica campestris (syn. rapa). *Genetics*, **153**(1), 391–400.

Swanson, SJ, Bethke, PC, & Jones, RL. 1998. Barley aleurone cells contain two types of vacuoles. Characterization Of lytic organelles by use of fluorescent probes. *THE PLANT CELL ONLINE*, **10**(5), 685–698.

Takayama, S, Shimosato, H, Shiba, H, Funato, M, Che, F S, Watanabe, M, Iwano, M, & Isogai, A. 2001. Direct ligand-receptor complex interaction controls Brassica self-incompatibility. *Nature*, **413**(6855), 534–538.

Tanaka, H, Onouchi, H, Kondo, M, Hara-Nishimura, I, Nishimura, M, Machida, C, & Machida, Y. 2001. A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. *Development (Cambridge, England)*, **128**(23), 4681–4689.

Tang, Weihua, Kelley, Dior, Ezcurra, Inés, Cotter, Robyn, & McCormick, Sheila. 2004. LeSTIG1, an extracellular binding partner for the pollen receptor kinases LePRK1 and LePRK2, promotes pollen tube growth in vitro. *The Plant journal : for cell and molecular biology*, **39**(3), 343–353.

Taylor, W R. 1997. Residual colours: a proposal for aminochromography. *Protein engineering*, **10**(7), 743–746.

Therien, Alex G Grant Fiona E M Deber Charles M. 2001. Interhelical hydrogen bonds in the CFTR membrane domain. *Nature Structural Biology*, **8**, 579–601.

Thompson, J D, Higgins, D G, & Gibson, T J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**(22), 4673–4680.

Thompson, R D, Hueros, G, Becker, H, & Maitz, M. 2001. Development and functions of seed transfer cells. *Plant Sci*, **160**(5), 775–783.

Tirlapur, U K, Kranz, E, & Cresti, M. 1995. Characterisation of isolated egg cells, in vitro fusion products and zygotes of Zea mays L. using the technique of image analysis and confocal laser scanning microscopy. *Zygote (Cambridge, England)*, **3**(1), 57–64.

Tsien, R Y. 1998. The green fluorescent protein. *Annual review of biochemistry*, **67**, 509–544.

Uebler, Suzanne, Dresselhaus, Thomas, & Márton, Mihaela-Luiza. 2013. Species-specific interaction of EA1 with the maize pollen tube apex. *Plant signaling & behavior*, **8**(10), doi: 10.4161–psb.25682.

Uversky, Vladimir N. 2002. Natively unfolded proteins: A point where biology waits for physics. *Protein Science*, **11**(4), 739–756.

Vallejo, Luis, & Rinas, Ursula. 2004. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories*, **3**(1), 11.

Van de Velde, Willem, Zehirov, Grigor, Szatmari, Agnes, Debreczeny, Monika, Ishihara, Hironobu, Kevei, Zoltan, Farkas, Attila, Mikulass, Kata, Nagy, Andrea, Tiricz,

Hilda, Satiat-Jeunemaitre, Béatrice, Alunni, Benoit, Bourge, Mickael, Kucho, Kenichi, Abe, Mikiko, Kereszt, Attila, Maroti, Gergely, Uchiumi, Toshiki, Kondorosi, Eva, & Mergaert, Peter. 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*, **327**(5969), 1122–1126.

Van Lammeren, A A M. 1986a. Developmental morphology and cytology of the young maize embryo (Zea mays L.). *Acta botanica neerlandica*, **35**(3), 169–188.

Van Lammeren, AAM. 1986b. Comparative ultrastructural study of the megagametophytes in two strains of Zea mays L. before and after fertilization. **86**, 1–37.

Vaucheret, H, & Fagard, M. 2001. Transcriptional gene silencing in plants: targets, inducers and regulators. *Trends in genetics : TIG*, **17**(1), 29–35.

Verdaguer, B, de Kochko, A, Fux, C I, Beachy, R N, & Fauquet, C. 1998. Functional organization of the cassava vein mosaic virus (CsVMV) promoter. *Plant molecular biology*, **37**(6), 1055–1067.

Voinnet, Olivier, Rivas, Susana, Mestre, Pere, & Baulcombe, David. 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *The Plant journal : for cell and molecular biology*, **33**(5), 949–956.

Walker, J M. 2005. *The proteomics protocols handbook*. Humana Press.

Wang, Dong, Griffitts, Joel, Starker, Colby, Fedorova, Elena, Limpens, Erik, Ivanov, Sergey, Bisseling, Ton, & Long, Sharon. 2010. A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science*, **327**(5969), 1126–1129.

Wang, Lu, & Ruan, Yong-Ling. 2012a. New Insights into Roles of Cell Wall Invertase in Early Seed Development Revealed by Comprehensive Spatial and Temporal Expression Patterns of GhCWIN1 in Cotton. *Plant Physiology*, **160**(2), 777–787.

Wang, Lu, & Ruan, Yong-Ling. 2012b. New insights into roles of cell wall invertase in early seed development revealed by comprehensive spatial and temporal expression patterns of GhCWIN1 in cotton. **160**, 777–787.

Wang, Lu, & Ruan, Yong-Ling. 2013. Regulation of cell division and expansion by sugar and auxin signaling. *Frontiers in plant science*, **4**, 163.

Wang, Zhong, Gerstein, Mark, & Snyder, Michael. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, **10**(1), 57–63.

Wassenegger, M, & Pélissier, T. 1998. A model for RNA-mediated gene silencing in higher plants. *Plant molecular biology*, **37**(2), 349–362.

Waterhouse, A M Procter J B Martin D M A Clamp M, & Barton, G J. 2009. Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**(9), 1189–1191.

Waters, Amanda J, Makarevitch, Irina, Eichten, Steve R, Swanson-Wagner, Ruth A, Yeh, Cheng-Ting, Xu, Wayne, Schnable, Patrick S, Vaughn, Matthew W, Gehring, Mary, & Springer, Nathan M. 2011. Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *THE PLANT CELL ONLINE*, **23**(12), 4221–4233.

Weschke, Winfriede, Panitz, Reinhard, Gubatz, Sabine, Wang, Qing, Radchuk, Ruslana, Weber, Hans, & Wobus, Ulrich. 2003. The role of invertases and hexose transporters in controlling sugar ratios in maternal and filial tissues of barley caryopses during early development. *The Plant journal : for cell and molecular biology*, **33**(2), 395–411.

Wheeler, Michael J, de Graaf, Barend H J, Hadjiosif, Natalie, Perry, Ruth M, Poulter, Natalie S, Osman, Kim, Vatovec, Sabina, Harper, Andrea, Franklin, F Christopher H, & Franklin-Tong, Vernonica E. 2009. Identification of the pollen self-incompatibility determinant in Papaver rhoeas. *Nature*, **459**(7249), 992–995.

Whitmore, Lee, & Wallace, B A. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers*, **89**(5), 392–400.

Wilhelm, Brian T, Marguerat, Samuel, Watt, Stephen, Schubert, Falk, Wood, Valerie, Goodhead, Ian, Penkett, Christopher J, Rogers, Jane, & Bähler, Jürg. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, **453**(7199), 1239–1243.

Williams, Joseph H, & Friedman, William E. 2002. Identification of diploid endosperm in an early angiosperm lineage. *Nature*, **415**(6871), 522–526.

Wilson, R C, & Doudna, J A. 2013. Molecular mechanisms of RNA interference. *Annual review of biophysics*, **42**, 217–239.

Winter, Debbie, Vinegar, Ben, Nahal, Hardeep, Ammar, Ron, Wilson, Greg V, & Provart, Nicholas J. 2007. An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One*, **2**(8), e718.

Wobus, U, & Weber, H. 1999. Sugars as signal molecules in plant seed development. *Biological chemistry*, **380**(7-8), 937–944.

Woriedh, Mayada, Wolf, Sebastian, Márton, Mihaela L, Hinze, Axel, Gahrtz, Manfred, Becker, Dirk, & Dresselhaus, Thomas. 2013. External application of gametophyte-specific ZmPMEI1 induces pollen tube burst in maize. *Plant reproduction*, **26**(3), 255–266.

Xiong, Yuqing, Mei, Wenbin, Kim, Eun-Deok, Mukherjee, Krishanu, Hassanein, Hatem, Barbazuk, William B, Sung, Sibum, Kolaczkowski, Bryan, & Kang, Byung-Ho. 2014. Adaptive expansion of the maize maternally expressed gene (Meg) family involves changes in expression patterns and protein secondary structures of its members. *BMC plant biology*, **14**(1), 204.

Yang, Weibing, Gao, Mingjun, Yin, Xin, Liu, Jiyun, Xu, Yonghan, Zeng, Longjun, Li, Qun, Zhang, Shubiao, Wang, Junmin, Zhang, Xiaoming, & He, Zuhua. 2013. Control of rice embryo development, shoot apical meristem maintenance, and grain yield by a novel cytochrome p450. *Molecular plant*, **6**(6), 1945–1960.

Yeung, E C, & Meinke, D W. 1993. Embryogenesis in angiosperms: development of the suspensor. *The Plant Cell*, **5**, 1371–1381.

Yeung, Edward C. 1980. Embryogeny of Phaseolus: the Role of the Suspensor. *Zeitschrift für Pflanzenphysiologie*, **96**(1), 17–28.

Young, Anna. 2007. Structural insights into the clathrin coat. *Seminars in cell & developmental biology*, **18**(4), 448–458.

Young, T E, & Gallie, D R. 2000. Programmed cell death during endosperm development. *Programmed Cell Death in Higher Plants*, **1**, 39–57.

Young, T E, Gallie, D R, & DeMason, D A. 1997. Ethylene-Mediated Programmed Cell Death during Maize Endosperm Development of Wild-Type and shrunken2 Genotypes. *Plant P*, **115**, 737–751.

Zheng, Yankun, & Wang, Zhong. 2010. Current opinions on endosperm transfer cells in maize. *Plant Cell Rep*, **29**(9), 935–942.

Zhu, S, Gao, B, & Tytgat, J. 2005. Phylogenetic distribution, functional epitopes and evolution of the CS$\alpha\beta$ superfamily. *Cellular and molecular life sciences : CMLS*, **62**(19–20), 2257–2269.